```
          *   207522   207621: gap of 100 bp
          *   207622   210083: contig of 2462 bp in length
          *   210084   210183: gap of 100 bp
          *   210184   211973: contig of 1790 bp in length.
FEATURES             Location/Qualifiers
     source          1. .211973
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q24"
                     /clone="RP11-820M9"
     misc_feature    1. .30856
                     /note="assembly_fragment"
     misc_feature    30957. .46182
                     /note="assembly_fragment clone_end:SP6 vector_side:right"
     misc_feature    46283. .60789
                     /note="assembly_fragment"
     misc_feature    60890. .73659
                     /note="assembly_fragment"
     misc_feature    73760. .88723
                     /note="assembly_fragment"
     misc_feature    88824. .93706
                     /note="assembly_fragment"
     misc_feature    93807. .103722
                     /note="assembly_fragment"
     misc_feature    103823. .112270
                     /note="assembly_fragment"
     misc_feature    112371. .119605
                     /note="assembly_fragment"
     misc_feature    119706. .124160
                     /note="assembly_fragment"
     misc_feature    124261. .129556
                     /note="assembly_fragment"
     misc_feature    129657. .134737
                     /note="assembly_fragment"
     misc_feature    134838. .139920
                     /note="assembly_fragment"
     misc_feature    140021. .145577
                     /note="assembly_fragment"
     misc_feature    145678. .149898
                     /note="assembly_fragment"
     misc_feature    149999. .154667
                     /note="assembly_fragment"

Alignment Scores:
Pred. No.:                3.78e-51      Length:          211973
Score:                    546.00        Matches:         107
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    99.07%        Mismatches:      0
Query Match:              99.82%        Indels:          0
DB:                       2             Gaps:            0

US-10-679-865A-2 (1-108) x AP001163 (1-211973)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50092 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCT 50151

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50152 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 50211

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50212 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 50271

Qy         61 LysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50272 AAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAAT 50331

Qy         81 ValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnValThrLeuLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50332 GTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTA 50391
```

```
Qy         101 CysAsnSerSerAlaGluAlaLeu 108
                ||||||||||||||||||||||||
Db      50392 TGCAATTCCTCTGCTGAGGCCCTG 50415

RESULT 7
AC009544
LOCUS       AC009544               212357 bp    DNA     linear   HTG 22-APR-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-85D24 map 11, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC009544
VERSION     AC009544.4  GI:7637232
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 212357)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-85D24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212357)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 22, 2000 this sequence version replaced gi:6454062.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1649
                Center clone name: 85_D_24
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 3% of reads
                Chemistry: Dye-terminator Big Dye; 97% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 166296 bases at least Q40
                Consensus quality: 188982 bases at least Q30
                Consensus quality: 201949 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 211057; sum-of-contigs
                Quality coverage: 4.9 in Q20 bases; agarose-fp
                Quality coverage: 4.1 in Q20 base.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     1174: contig of 1174 bp in length
```

This Page Blank (uspto)

```
            *     1175     1274: gap of 100 bp
            *     1275     2941: contig of 1667 bp in length
            *     2942     3041: gap of 100 bp
            *     3042     5780: contig of 2739 bp in length
            *     5781     5880: gap of 100 bp
            *     5881    12941: contig of 7061 bp in length
            *    12942    13041: gap of 100 bp
            *    13042    23383: contig of 10342 bp in length
            *    23384    23483: gap of 100 bp
            *    23484    32867: contig of 9384 bp in length
            *    32868    32967: gap of 100 bp
            *    32968    43838: contig of 10871 bp in length
            *    43839    43938: gap of 100 bp
            *    43939    55437: contig of 11499 bp in length
            *    55438    55537: gap of 100 bp
            *    55538    71158: contig of 15621 bp in length
            *    71159    71258: gap of 100 bp
            *    71259    87581: contig of 16323 bp in length
            *    87582    87681: gap of 100 bp
            *    87682   105529: contig of 17848 bp in length
            *   105530   105629: gap of 100 bp
            *   105630   120782: contig of 15153 bp in length
            *   120783   120882: gap of 100 bp
            *   120883   161108: contig of 40226 bp in length
            *   161109   161208: gap of 100 bp
            *   161209   212357: contig of 51149 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212357
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-85D24"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1174
                     /note="assembly_fragment"
     misc_feature    1275. .2941
                     /note="assembly_fragment"
     misc_feature    3042. .5780
                     /note="assembly_fragment"
     misc_feature    5881. .12941
                     /note="assembly_fragment"
     misc_feature    13042. .23383
                     /note="assembly_fragment"
     misc_feature    23484. .32867
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    32968. .43838
                     /note="assembly_fragment"
     misc_feature    43939. .55437
                     /note="assembly_fragment"
     misc_feature    55538. .71158
                     /note="assembly_fragment"
     misc_feature    71259. .87581
                     /note="assembly_fragment"
     misc_feature    87682. .105529
                     /note="assembly_fragment"
     misc_feature    105630. .120782
                     /note="assembly_fragment"
     misc_feature    120883. .161108
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    161209. .212357
                     /note="assembly_fragment"
ORIGIN

Alignment Scores:
Pred. No.:              3.79e-51        Length:         212357
Score:                  546.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-10-679-865A-2 (1-108) x AC009544 (1-212357)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db      63790 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCT 63849

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63850 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 63909

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63910 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 63969

Qy         61 LysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63970 AAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAAT 64029

Qy         81 ValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnValThrLeuLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64030 GTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTA 64089

Qy        101 CysAsnSerSerAlaGluAlaLeu 108
              |||||||||||||||||||||||||
Db      64090 TGCAATTCCTCTGCTGAGGCCCTG 64113

RESULT 8
AP001359/c
LOCUS       AP001359               221535 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q clone:RP11-820L6, complete
            sequences.
ACCESSION   AP001359
VERSION     AP001359.4  GI:14189751
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 221535)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On May 22, 2001 this sequence version replaced gi:9857628.
FEATURES             Location/Qualifiers
     source          1. .221535
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-820L6"
ORIGIN

Alignment Scores:
Pred. No.:              3.96e-51        Length:         221535
Score:                  546.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
```

This Page Blank (uspto)

```
     source          1. .211973
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q24"
                     /clone="RP11-820M9"
     misc_feature    1. .30856
                     /note="assembly_fragment"
     misc_feature    30957. .46182
                     /note="assembly_fragment clone_end:SP6 vector_side:right"
     misc_feature    46283. .60789
                     /note="assembly_fragment"
     misc_feature    60890. .73659
                     /note="assembly_fragment"
     misc_feature    73760. .83723
                     /note="assembly_fragment"
     misc_feature    83824. .93706
                     /note="assembly_fragment"
     misc_feature    93807. .103722
                     /note="assembly_fragment"
     misc_feature    103823. .112270
                     /note="assembly_fragment"
     misc_feature    112371. .119605
                     /note="assembly_fragment"
     misc_feature    119706. .124160
                     /note="assembly_fragment"
     misc_feature    124261. .129556
                     /note="assembly_fragment"
     misc_feature    129657. .134737
                     /note="assembly_fragment"
     misc_feature    134838. .139920
                     /note="assembly_fragment"
     misc_feature    140021. .145577
                     /note="assembly_fragment"
     misc_feature    145678. .149898
                     /note="assembly_fragment"
     misc_feature    149999. .154667

  Query Match             99.8%;  Score 860.4;  DB 2;  Length 211973;
  Best Local Similarity   99.9%;  Pred. No. 3.9e-209;
  Matches  861;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49799 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 49858

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49859 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 49918

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49919 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGCATCCCACCACTTTTTTCTTCCCATC 49978

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49979 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 50038

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50039 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 50098

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50099 CTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCTGAGTGTG 50158

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50159 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 50218

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50219 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 50278

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50279 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 50338

Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50339 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 50398

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50399 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 50458

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50459 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 50518

Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50519 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 50578

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50579 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 50638

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||||
Db      50639 TTTAATTAAAACAAATCTTGCA 50660
```

RESULT 7
AC009544

```
LOCUS       AC009544              212357 bp    DNA     linear   HTG 22-APR-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-85D24 map 11, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC009544
VERSION     AC009544.4  GI:7637232
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 212357)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-85D24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212357)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 22, 2000 this sequence version replaced gi:6454062.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
```

This Page Blank (uspto)

```
        Web site: http://www-seq.wi.mit.edu
        Contact: sequence_submissions@genome.wi.mit.edu
    ---------- Project Information
        Center project name: L1649
        Center clone name: 85_D_24
    ---------- Summary Statistics
        Sequencing vector: M13; M77815; 100% of reads
        Chemistry: Dye-primer-amersham; 3% of reads
        Chemistry: Dye-terminator Big Dye; 97% of reads
        Assembly program: Phrap; version 0.960731
        Consensus quality: 166296 bases at least Q40
        Consensus quality: 188982 bases at least Q30
        Consensus quality: 201949 bases at least Q20
        Insert size: 176000; agarose-fp
        Insert size: 211057; sum-of-contigs
        Quality coverage: 4.9 in Q20 bases; agarose-fp
        Quality coverage: 4.1 in Q20 base.
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 14 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
    * arbitrary. Gaps between the contigs are represented as
    * runs of N, but the exact sizes of the gaps are unknown.
    * This record will be updated with the finished sequence
    * as soon as it is available and the accession number will
    * be preserved.
    *        1     1174: contig of 1174 bp in length
    *     1175     1274: gap of 100 bp
    *     1275     2941: contig of 1667 bp in length
    *     2942     3041: gap of 100 bp
    *     3042     5780: contig of 2739 bp in length
    *     5781     5880: gap of 100 bp
    *     5881    12941: contig of 7061 bp in length
    *    12942    13041: gap of 100 bp
    *    13042    23383: contig of 10342 bp in length
    *    23384    23483: gap of 100 bp
    *    23484    32867: contig of 9384 bp in length
    *    32868    32967: gap of 100 bp
    *    32968    43838: contig of 10871 bp in length
    *    43839    43938: gap of 100 bp
    *    43939    55437: contig of 11499 bp in length
    *    55438    55537: gap of 100 bp
    *    55538    71158: contig of 15621 bp in length
    *    71159    71258: gap of 100 bp
    *    71259    87581: contig of 16323 bp in length
    *    87582    87681: gap of 100 bp
    *    87682   105529: contig of 17848 bp in length
    *   105530   105629: gap of 100 bp
    *   105630   120782: contig of 15153 bp in length
    *   120783   120882: gap of 100 bp
    *   120883   161108: contig of 40226 bp in length
    *   161109   161208: gap of 100 bp
    *   161209   212357: contig of 51149 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212357
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-85D24"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1174
                     /note="assembly_fragment"
     misc_feature    1275. .2941
                     /note="assembly_fragment"
     misc_feature    3042. .5780
                     /note="assembly_fragment"
     misc_feature    5881. .12941
                     /note="assembly_fragment"
     misc_feature    13042. .23383
                     /note="assembly_fragment"
     misc_feature    23484. .32867
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    32968. .43838
                     /note="assembly_fragment"
     misc_feature    43939. .55437
                     /note="assembly_fragment"
     misc_feature    55538. .71158
                     /note="assembly_fragment"
     misc_feature    71259. .87581
                     /note="assembly_fragment"
     misc_feature    87682. .105529
                     /note="assembly_fragment"
     misc_feature    105630. .120782
                     /note="assembly_fragment"
     misc_feature    120883. .161108
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    161209. .212357
                     /note="assembly_fragment"
ORIGIN

  Query Match           99.8%;  Score 860.4;  DB 2;  Length 212357;
  Best Local Similarity  99.9%;  Pred. No. 3.9e-209;
  Matches  861;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63497 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 63556

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63557 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 63616

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63617 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 63676

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63677 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 63736

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63737 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 63796

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db      63797 CTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCTGAGTGTG 63856

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63857 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 63916

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63917 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 63976

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63977 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 64036

Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64037 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 64096

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64097 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 64156

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64157 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 64216
```

This Page Blank (uspto)

```
Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64217 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 64276

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64277 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 64336

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||
Db      64337 TTTAATTAAAACAAATCTTGCA 64358
```

```
RESULT 8
AP001359/c
LOCUS       AP001359               221535 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q clone:RP11-820L6, complete
            sequences.
ACCESSION   AP001359
VERSION     AP001359.4  GI:14189751
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 221535)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On May 22, 2001 this sequence version replaced gi:9857628.
FEATURES             Location/Qualifiers
     source          1. .221535
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-820L6"
ORIGIN
```

```
  Query Match           99.8%;  Score 860.4;  DB 9;  Length 221535;
  Best Local Similarity  99.9%;  Pred. No. 3.9e-209;
  Matches  861;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84173 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 84114

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84113 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 84054

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84053 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 83994

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83993 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 83934

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      83933 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 83874

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83873 CTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCTGAGTGTG 83814

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83813 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 83754

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83753 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 83694

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83693 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 83634

Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83633 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 83574

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83573 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 83514

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83513 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 83454

Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83453 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 83394

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83393 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 83334

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||
Db      83333 TTTAATTAAAACAAATCTTGCA 83312
```

```
RESULT 9
AC016859
LOCUS       AC016859               79544 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-13K20, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016859
VERSION     AC016859.2  GI:7144864
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 79544)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-13K20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 79544)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Gallagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 31, 2005, 11:24:30 ; Search time 7670 Seconds
(without alignments)
682.290 Million cell updates/sec

Title: US-10-679-865A-2
Perfect score: 547
Sequence: 1 MVTLLPIEGQEVHFFEILES.....PSYLTRIQVTLLCNSSAEAL 108

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US10679865/runat_30082005_150801_15911/app_query.fasta_1.263
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10679865_@CGN_1_1_3731_@runat_30082005_150801_15911 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		547	100.0	862	9	AF303179	AF303179 Homo sapi
c 2		546	99.8	102807	2	AP000674	AP000674 Homo sapi
3		546	99.8	119419	2	AP000666	AP000666 Homo sapi
c 4		546	99.8	164784	2	AP001384	AP001384 Homo sapi
c 5		546	99.8	190360	9	AP001924	AP001924 Homo sapi
6		546	99.8	211973	2	AP001163	AP001163 Homo sapi
7		546	99.8	212357	2	AC009544	AC009544 Homo sapi
c 8		546	99.8	221535	9	AP001359	AP001359 Homo sapi
9		346	63.3	1572	6	CQ738864	CQ738864 Sequence
10		201	36.7	79544	2	AC016859	AC016859 Homo sapi
11		86.5	15.8	300349	1	AE017319	AE017319 Desulfovi
c 12		85	15.5	1384	10	MMU278829	AJ278829 Mus muscu
c 13		85	15.5	2161	10	BC051938	BC051938 Mus muscu
c 14		85	15.5	2170	10	BC043041	BC043041 Mus muscu
15		85	15.5	185630	2	AC074154	AC074154 Mus muscu
16		85	15.5	269576	2	AC073684	AC073684 Mus muscu
17		83	15.2	111159	9	AC008271	AC008271 Homo sapi
c 18		82.5	15.1	168243	9	AC008526	AC008526 Homo sapi
c 19		82	15.0	93821	9	AC004254	AC004254 Homo sapi
20		82	15.0	183102	5	BX548069	BX548069 Zebrafish
c 21		82	15.0	250119	2	AC135607	AC135607 Pan trogl
c 22		81.5	14.9	110000	2	AC118787_5	Continuation (6 of
c 23		81.5	14.9	136889	2	AC135296	AC135296 Rattus no
24		81.5	14.9	162199	2	AC102841	AC102841 Mus muscu
c 25		81.5	14.9	196321	2	AC137253	AC137253 Rattus no
c 26		81	14.8	174283	2	AC013723	AC013723 Homo sapi
27		81	14.8	199827	2	AC026373	AC026373 Homo sapi
28		81	14.8	202383	9	AC009973	AC009973 Homo sapi
c 29		80	14.6	114051	8	OSJN00036	AL606598 Oryza sat
c 30		80	14.6	221782	2	AC104540	AC104540 Mus muscu
31		79	14.4	237384	2	AC131424	AC131424 Rattus no
32		79	14.4	250870	2	AC112584	AC112584 Rattus no
c 33		79	14.4	259474	2	AC095692	AC095692 Rattus no
c 34		79	14.4	260989	2	AC131615	AC131615 Rattus no
c 35		78.5	14.4	2449	10	BC005722	BC005722 Mus muscu
c 36		78.5	14.4	3480	10	AB045293	AB045293 Mus muscu
c 37		78.5	14.4	26042	6	AX695356	AX695356 Sequence
c 38		78.5	14.4	64148	2	AC101402	AC101402 Mus muscu
c 39		78.5	14.4	167859	2	CR354378	CR354378 Danio rer
c 40		78.5	14.4	173982	5	BX323052	BX323052 Zebrafish
c 41		78.5	14.4	174678	9	AL450332	AL450332 Human DNA
42		78.5	14.4	177041	10	AL670958	AL670958 Mouse DNA
c 43		78.5	14.4	186632	10	AC140047	AC140047 Mus muscu
c 44		78.5	14.4	286283	10	AC139214	AC139214 Mus muscu
45		78	14.3	168544	2	AC131134	AC131134 Rattus no

ALIGNMENTS

RESULT 1
AF303179
LOCUS AF303179 862 bp mRNA linear PRI 14-JUN-2004
DEFINITION Homo sapiens BRCC2 mRNA, complete cds.
ACCESSION AF303179
VERSION AF303179.1 GI:23393067
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 862)
AUTHORS Broustas,C.G., Gokhale,P.C., Rahman,A., Dritschilo,A., Ahmad,I. and Kasid,U.
TITLE BRCC2, a Novel BH3-like-domain-containing protein, induces apoptosis in a caspase-dependent manner
JOURNAL J. Biol. Chem. 279 (25), 26780-26788 (2004)
PUBMED 15069058
REFERENCE 2 (bases 1 to 862)
AUTHORS Gokhale,P.C., Kumar,D., Monia,B.P., Rahman,A., Dritschilo,A. and Kasid,U.
TITLE Novel Transcript in Human Breast Cancer Cells
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 862)
AUTHORS Gokhale,P.C., Rahman,A., Dritschilo,A. and Kasid,U.
TITLE Direct Submission
JOURNAL Submitted (05-SEP-2000) Radiation Medicine, Georgetown University, 3970 Reservoir Rd, NW, Washington, DC 20007, USA

```
FEATURES             Location/Qualifiers
     source          1. .862
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /tissue_type="breast cancer"
     CDS             294. .620
                     /codon_start=1
                     /product="BRCC2"
                     /protein_id="AAN31127.1"
                     /db_xref="GI:23393068"
                     /translation="MVTLLPIEGQEVHFFEILESECVLYTGWIERASGSSIYPEAKAR
                     LPLEALLGSNKEPMLPKETVLSLKRYNLGSSAMKRNVPGHVLQRPSYLTRIQVTLLCN
                     SSAEAL"
ORIGIN

Alignment Scores:
Pred. No.:              7.22e-54        Length:         862
Score:                  547.00          Matches:        108
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-10-679-865A-2 (1-108) x AF303179 (1-862)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCT 353

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        354 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 413

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        414 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 473

Qy         61 LysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        474 AAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAAT 533

Qy         81 ValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnValThrLeuLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        534 GTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTA 593

Qy        101 CysAsnSerSerAlaGluAlaLeu 108
              |||||||||||||||||||||||||
Db        594 TGCAATTCCTCTGCTGAGGCCCTG 617

RESULT 2
AP000674/c
LOCUS       AP000674               102807 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone CMB9-8G24 map 11q24, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AP000674
VERSION     AP000674.2  GI:8118863
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 102807)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 102,807 genomic DNA of 11q24
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 102807)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997548.
            ---------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            ---------- Project Information
                Center project name: HumDraft11
                Center clone name: CMB9-8G24
            ---------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 90471 bases at least Q40
                Consensus quality: 96773 bases at least Q30
                Consensus quality: 99873 bases at least Q20
                Insert size: 101207; sum-of-contigs
                Quality coverage: 4.21x in Q20 bases; sum-of-contigs
            ----------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            17 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                    1     23670 contig of  23670 bp in length
                23771     39763 contig of  15993 bp in length
                39864     50918 contig of  11055 bp in length
                51019     61370 contig of  10352 bp in length
                61471     70516 contig of   9046 bp in length
                70617     75763 contig of   5147 bp in length
                75864     81132 contig of   5269 bp in length
                81233     84386 contig of   3154 bp in length
                84487     87524 contig of   3038 bp in length
                87625     89403 contig of   1779 bp in length
                89504     91991 contig of   2488 bp in length
                92092     94299 contig of   2208 bp in length
                94400     96329 contig of   1930 bp in length
                96430     97921 contig of   1492 bp in length
                98022    100412 contig of   2391 bp in length
               100513    101594 contig of   1082 bp in length
               101695    102807 contig of   1113 bp in length
            Sequence updated (26-May-2000).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    23670: contig of 23670 bp in length
            *   23671    23770: gap of 100 bp
            *   23771    39763: contig of 15993 bp in length
            *   39764    39863: gap of 100 bp
            *   39864    50918: contig of 11055 bp in length
            *   50919    51018: gap of 100 bp
            *   51019    61370: contig of 10352 bp in length
            *   61371    61470: gap of 100 bp
            *   61471    70516: contig of 9046 bp in length
            *   70517    70616: gap of 100 bp
            *   70617    75763: contig of 5147 bp in length
            *   75764    75863: gap of 100 bp
            *   75864    81132: contig of 5269 bp in length
            *   81133    81232: gap of 100 bp
            *   81233    84386: contig of 3154 bp in length
            *   84387    84486: gap of 100 bp
```

```
            *    84487    87524: contig of 3038 bp in length
            *    87525    87624: gap of 100 bp
            *    87625    89403: contig of 1779 bp in length
            *    89404    89503: gap of 100 bp
            *    89504    91991: contig of 2488 bp in length
            *    91992    92091: gap of 100 bp
            *    92092    94299: contig of 2208 bp in length
            *    94300    94399: gap of 100 bp
            *    94400    96329: contig of 1930 bp in length
            *    96330    96429: gap of 100 bp
            *    96430    97921: contig of 1492 bp in length
            *    97922    98021: gap of 100 bp
            *    98022   100412: contig of 2391 bp in length
            *   100413   100512: gap of 100 bp
            *   100513   101594: contig of 1082 bp in length
            *   101595   101694: gap of 100 bp
            *   101695   102807: contig of 1113 bp in length.
FEATURES             Location/Qualifiers
     source          1. .102807
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q24"
                     /clone="CMB9-8G24"
     misc_feature    1. .23670
                     /note="assembly_fragment"
     misc_feature    23771. .39763
                     /note="assembly_fragment"
     misc_feature    39864. .50918
                     /note="assembly_fragment"
     misc_feature    51019. .61370
                     /note="assembly_fragment"
     misc_feature    61471. .70516
                     /note="assembly_fragment"
     misc_feature    70617. .75763
                     /note="assembly_fragment"
     misc_feature    75864. .81132
                     /note="assembly_fragment"
     misc_feature    81233. .84386
                     /note="assembly_fragment"
     misc_feature    84487. .87524
                     /note="assembly_fragment"
     misc_feature    87625. .89403
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    89504. .91991
                     /note="assembly_fragment"
     misc_feature    92092. .94299
                     /note="assembly_fragment"
     misc_feature    94400. .96329
                     /note="assembly_fragment"
     misc_feature    96430. .97921
                     /note="assembly_fragment"
     misc_feature    98022. .100412
                     /note="assembly_fragment"
     misc_feature    100513. .101594
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    101695. .102807
                     /note="assembly_fragment"
ORIGIN

Alignment Scores:
Pred. No.:              1.72e-51        Length:         102807
Score:                  546.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-10-679-865A-2 (1-108) x AP000674 (1-102807)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
```

```
Db      31628 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCT 31569

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31568 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 31509

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31508 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 31449

Qy         61 LysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31448 AAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAAT 31389

Qy         81 ValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnValThrLeuLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31388 GTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTA 31329

Qy        101 CysAsnSerSerAlaGluAlaLeu 108
              |||||||||||||||||||||||
Db      31328 TGCAATTCCTCTGCTGAGGCCCTG 31305
```

```
RESULT 3
AP000666
LOCUS       AP000666               119419 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone CMB9-53P8 map 11q23, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AP000666
VERSION     AP000666.2  GI:8118855
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 119419)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 119,419 genomic DNA of 11q23
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 119419)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997540.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: CMB9-53P8
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 109205 bases at least Q40
                Consensus quality: 113734 bases at least Q30
                Consensus quality: 116089 bases at least Q20
                Insert size: 117519; sum-of-contigs
                Quality coverage: 4.59x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            20 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
```

```
        as soon as it is available and the accession number will be
        preserved
             1     22810 contig of  22810 bp in length
         22911     38820 contig of  15910 bp in length
         38921     52522 contig of  13602 bp in length
         52623     62588 contig of   9966 bp in length
         62689     72547 contig of   9859 bp in length
         72648     78455 contig of   5808 bp in length
         78556     83981 contig of   5426 bp in length
         84082     89379 contig of   5298 bp in length
         89480     93708 contig of   4229 bp in length
         93809     96503 contig of   2695 bp in length
         96604    100510 contig of   3907 bp in length
        100611    104144 contig of   3534 bp in length
        104245    107321 contig of   3077 bp in length
        107422    109944 contig of   2523 bp in length
        110045    111930 contig of   1886 bp in length
        112031    113584 contig of   1554 bp in length
        113685    115373 contig of   1689 bp in length
        115474    117116 contig of   1643 bp in length
        117217    118797 contig of   1581 bp in length
        118898    119419 contig of    522 bp in length
        Sequence updated (26-May-2000).
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 20 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *        1     22810: contig of 22810 bp in length
        *    22811     22910: gap of 100 bp
        *    22911     38820: contig of 15910 bp in length
        *    38821     38920: gap of 100 bp
        *    38921     52522: contig of 13602 bp in length
        *    52523     52622: gap of 100 bp
        *    52623     62588: contig of 9966 bp in length
        *    62589     62688: gap of 100 bp
        *    62689     72547: contig of 9859 bp in length
        *    72548     72647: gap of 100 bp
        *    72648     78455: contig of 5808 bp in length
        *    78456     78555: gap of 100 bp
        *    78556     83981: contig of 5426 bp in length
        *    83982     84081: gap of 100 bp
        *    84082     89379: contig of 5298 bp in length
        *    89380     89479: gap of 100 bp
        *    89480     93708: contig of 4229 bp in length
        *    93709     93808: gap of 100 bp
        *    93809     96503: contig of 2695 bp in length
        *    96504     96603: gap of 100 bp
        *    96604    100510: contig of 3907 bp in length
        *   100511    100610: gap of 100 bp
        *   100611    104144: contig of 3534 bp in length
        *   104145    104244: gap of 100 bp
        *   104245    107321: contig of 3077 bp in length
        *   107322    107421: gap of 100 bp
        *   107422    109944: contig of 2523 bp in length
        *   109945    110044: gap of 100 bp
        *   110045    111930: contig of 1886 bp in length
        *   111931    112030: gap of 100 bp
        *   112031    113584: contig of 1554 bp in length
        *   113585    113684: gap of 100 bp
        *   113685    115373: contig of 1689 bp in length
        *   115374    115473: gap of 100 bp
        *   115474    117116: contig of 1643 bp in length
        *   117117    117216: gap of 100 bp
        *   117217    118797: contig of 1581 bp in length
        *   118798    118897: gap of 100 bp
        *   118898    119419: contig of 522 bp in length.
FEATURES             Location/Qualifiers
     source          1. .119419
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q23"
                     /clone="CMB9-53P8"
          misc_feature    1. .22810
                     /note="assembly_fragment"
          misc_feature    22911. .38820
                     /note="assembly_fragment clone_end:T7 vector_side:right"
          misc_feature    38921. .52522
                     /note="assembly_fragment"
          misc_feature    52623. .62588
                     /note="assembly_fragment"
          misc_feature    62689. .72547
                     /note="assembly_fragment"
          misc_feature    72648. .78455
                     /note="assembly_fragment"
          misc_feature    78556. .83981
                     /note="assembly_fragment"
          misc_feature    84082. .89379
                     /note="assembly_fragment"
          misc_feature    89480. .93708
                     /note="assembly_fragment"
          misc_feature    93809. .96503
                     /note="assembly_fragment"
          misc_feature    96604. .100510
                     /note="assembly_fragment"
          misc_feature    100611. .104144
                     /note="assembly_fragment"
          misc_feature    104245. .107321
                     /note="assembly_fragment"
          misc_feature    107422. .109944
                     /note="assembly_fragment"
          misc_feature    110045. .111930
                     /note="assembly_fragment"
          misc_feature    112031. .113584
                     /note="assembly_fragment"
          misc_feature    113685. .115373
                     /note="assembly_fragment"
          misc_feature    115474. .117116
                     /note="assembly_fragment"
          misc_feature    117217. .118797
                     /note="assembly_fragment"
          misc_feature    118898. .119419
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
ORIGIN

Alignment Scores:
Pred. No.:              2.02e-51        Length:         119419
Score:                  546.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-10-679-865A-2 (1-108) x AP000666 (1-119419)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db      74653 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCT 74712

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74713 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 74772

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74773 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 74832

Qy         61 LysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74833 AAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAAT 74892
```

```
Qy          81 ValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnValThrLeuLeu 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       74893 GTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTA 74952

Qy         101 CysAsnSerSerAlaGluAlaLeu 108
               |||||||||||||||||||||||
Db       74953 TGCAATTCCTCTGCTGAGGCCCTG 74976

RESULT 4
AP001384/c
LOCUS       AP001384               164784 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-382J20 map 11q24, WORKING
            DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION   AP001384
VERSION     AP001384.2  GI:8117296
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164784)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 164,784 genomic DNA of 11q24
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 164784)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7243295.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: RP11-382J20
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 147339 bases at least Q40
                Consensus quality: 156937 bases at least Q30
                Consensus quality: 160723 bases at least Q20
                Insert size: 162984; sum-of-contigs
                Quality coverage: 4.61x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            19 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     29112 contig of  29112 bp in length
              29213     60097 contig of  30885 bp in length
              60198     74677 contig of  14480 bp in length
              74778     90573 contig of  15796 bp in length
              90674     99271 contig of   8598 bp in length
              99372    108087 contig of   8716 bp in length
             108188    116782 contig of   8595 bp in length
             116883    124764 contig of   7882 bp in length
             124865    130409 contig of   5545 bp in length
             130510    137878 contig of   7369 bp in length
             137979    143242 contig of   5264 bp in length
             143343    148670 contig of   5328 bp in length
             148771    151308 contig of   2538 bp in length
             151409    154990 contig of   3582 bp in length
             155091    157252 contig of   2162 bp in length
             157353    159636 contig of   2284 bp in length
             159737    161763 contig of   2027 bp in length
             161864    163588 contig of   1725 bp in length
             163689    164784 contig of   1096 bp in length
            Sequence updated (26-May-2000).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    29112: contig of 29112 bp in length
            *   29113    29212: gap of 100 bp
            *   29213    60097: contig of 30885 bp in length
            *   60098    60197: gap of 100 bp
            *   60198    74677: contig of 14480 bp in length
            *   74678    74777: gap of 100 bp
            *   74778    90573: contig of 15796 bp in length
            *   90574    90673: gap of 100 bp
            *   90674    99271: contig of 8598 bp in length
            *   99272    99371: gap of 100 bp
            *   99372   108087: contig of 8716 bp in length
            *  108088   108187: gap of 100 bp
            *  108188   116782: contig of 8595 bp in length
            *  116783   116882: gap of 100 bp
            *  116883   124764: contig of 7882 bp in length
            *  124765   124864: gap of 100 bp
            *  124865   130409: contig of 5545 bp in length
            *  130410   130509: gap of 100 bp
            *  130510   137878: contig of 7369 bp in length
            *  137879   137978: gap of 100 bp
            *  137979   143242: contig of 5264 bp in length
            *  143243   143342: gap of 100 bp
            *  143343   148670: contig of 5328 bp in length
            *  148671   148770: gap of 100 bp
            *  148771   151308: contig of 2538 bp in length
            *  151309   151408: gap of 100 bp
            *  151409   154990: contig of 3582 bp in length
            *  154991   155090: gap of 100 bp
            *  155091   157252: contig of 2162 bp in length
            *  157253   157352: gap of 100 bp
            *  157353   159636: contig of 2284 bp in length
            *  159637   159736: gap of 100 bp
            *  159737   161763: contig of 2027 bp in length
            *  161764   161863: gap of 100 bp
            *  161864   163588: contig of 1725 bp in length
            *  163589   163688: gap of 100 bp
            *  163689   164784: contig of 1096 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164784
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q24"
                     /clone="RP11-382J20"
     misc_feature    1. .29112
                     /note="assembly_fragment"
     misc_feature    29213. .60097
                     /note="assembly_fragment"
     misc_feature    60198. .74677
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    74778. .90573
                     /note="assembly_fragment"
     misc_feature    90674. .99271
                     /note="assembly_fragment"
     misc_feature    99372. .108087
                     /note="assembly_fragment"
```

```
     misc_feature    108188. .116782
                     /note="assembly_fragment"
     misc_feature    116883. .124764
                     /note="assembly_fragment"
     misc_feature    124865. .130409
                     /note="assembly_fragment"
     misc_feature    130510. .137878
                     /note="assembly_fragment"
     misc_feature    137979. .143242
                     /note="assembly_fragment"
     misc_feature    143343. .148670
                     /note="assembly_fragment"
     misc_feature    148771. .151308
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    151409. .154990
                     /note="assembly_fragment"
     misc_feature    155091. .157252
                     /note="assembly_fragment"
     misc_feature    157353. .159636
                     /note="assembly_fragment"
     misc_feature    159737. .161763
                     /note="assembly_fragment"
     misc_feature    161864. .163588
                     /note="assembly_fragment"
     misc_feature    163689. .164784
                     /note="assembly_fragment"
ORIGIN

Alignment Scores:
Pred. No.:              2.87e-51        Length:         164784
Score:                  546.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-10-679-865A-2 (1-108) x AP001384 (1-164784)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db      20751 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCT 20692

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20691 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 20632

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20631 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 20572

Qy         61 LysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20571 AAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAAT 20512

Qy         81 ValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnValThrLeuLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20511 GTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTA 20452

Qy        101 CysAsnSerSerAlaGluAlaLeu 108
              |||||||||||||||||||||||||
Db      20451 TGCAATTCCTCTGCTGAGGCCCTG 20428

RESULT 5
AP001924/c
LOCUS       AP001924              190360 bp    DNA     linear   PRI 01-FEB-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11 clone:RP11-166D19, complete
            sequence.
ACCESSION   AP001924
VERSION     AP001924.5  GI:28189518
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 190360)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Jan 31, 2003 this sequence version replaced gi:19879815.
FEATURES             Location/Qualifiers
     source          1. .190360
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-166D19"
ORIGIN

Alignment Scores:
Pred. No.:              3.36e-51        Length:         190360
Score:                  546.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     9               Gaps:           0

US-10-679-865A-2 (1-108) x AP001924 (1-190360)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db     165520 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCT 165461

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165460 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 165401

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165400 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 165341

Qy         61 LysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165340 AAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAAT 165281

Qy         81 ValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnValThrLeuLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165280 GTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTA 165221

Qy        101 CysAsnSerSerAlaGluAlaLeu 108
              |||||||||||||||||||||||||
Db     165220 TGCAATTCCTCTGCTGAGGCCCTG 165197

RESULT 6
AP001163
LOCUS       AP001163              211973 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-820M9 map 11q24, WORKING
            DRAFT SEQUENCE, 34 unordered pieces.
ACCESSION   AP001163
VERSION     AP001163.2  GI:8118731
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 211973)
```

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 211,973 genomic DNA of 11q24
JOURNAL Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 211973)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT On May 31, 2000 this sequence version replaced gi:6997849.
---------- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
---------- Project Information
Center project name: HumDraft11
Center clone name: RP11-820M9
---------- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 185228 bases at least Q40
Consensus quality: 199175 bases at least Q30
Consensus quality: 205608 bases at least Q20
Insert size: 208673; sum-of-contigs
Quality coverage: 4.21x in Q20 bases; sum-of-contigs
----------.
NOTE: This is a 'working draft' sequence. It currently consists of 34 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1	30856	contig of	30856	bp in length
30957	46182	contig of	15226	bp in length
46283	60789	contig of	14507	bp in length
60890	73659	contig of	12770	bp in length
73760	83723	contig of	9964	bp in length
83824	93706	contig of	9883	bp in length
93807	103722	contig of	9916	bp in length
103823	112270	contig of	8448	bp in length
112371	119605	contig of	7235	bp in length
119706	124160	contig of	4455	bp in length
124261	129556	contig of	5296	bp in length
129657	134737	contig of	5081	bp in length
134838	139920	contig of	5083	bp in length
140021	145577	contig of	5557	bp in length
145678	149898	contig of	4221	bp in length
149999	154667	contig of	4669	bp in length
154768	158909	contig of	4142	bp in length
159010	163218	contig of	4209	bp in length
163319	167426	contig of	4108	bp in length
167527	172410	contig of	4884	bp in length
172511	176689	contig of	4179	bp in length
176790	180030	contig of	3241	bp in length
180131	185024	contig of	4894	bp in length
185125	188834	contig of	3710	bp in length
188935	191247	contig of	2313	bp in length
191348	194377	contig of	3030	bp in length
194478	196487	contig of	2010	bp in length
196588	199459	contig of	2872	bp in length
199560	201207	contig of	1648	bp in length
201308	203675	contig of	2368	bp in length
203776	205300	contig of	1525	bp in length
205401	207521	contig of	2121	bp in length
207622	210083	contig of	2462	bp in length
210184	211973	contig of	1790	bp in length

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1	30856	contig of 30856 bp in length
* 30857	30956	gap of 100 bp
* 30957	46182	contig of 15226 bp in length
* 46183	46282	gap of 100 bp
* 46283	60789	contig of 14507 bp in length
* 60790	60889	gap of 100 bp
* 60890	73659	contig of 12770 bp in length
* 73660	73759	gap of 100 bp
* 73760	83723	contig of 9964 bp in length
* 83724	83823	gap of 100 bp
* 83824	93706	contig of 9883 bp in length
* 93707	93806	gap of 100 bp
* 93807	103722	contig of 9916 bp in length
* 103723	103822	gap of 100 bp
* 103823	112270	contig of 8448 bp in length
* 112271	112370	gap of 100 bp
* 112371	119605	contig of 7235 bp in length
* 119606	119705	gap of 100 bp
* 119706	124160	contig of 4455 bp in length
* 124161	124260	gap of 100 bp
* 124261	129556	contig of 5296 bp in length
* 129557	129656	gap of 100 bp
* 129657	134737	contig of 5081 bp in length
* 134738	134837	gap of 100 bp
* 134838	139920	contig of 5083 bp in length
* 139921	140020	gap of 100 bp
* 140021	145577	contig of 5557 bp in length
* 145578	145677	gap of 100 bp
* 145678	149898	contig of 4221 bp in length
* 149899	149998	gap of 100 bp
* 149999	154667	contig of 4669 bp in length
* 154668	154767	gap of 100 bp
* 154768	158909	contig of 4142 bp in length
* 158910	159009	gap of 100 bp
* 159010	163218	contig of 4209 bp in length
* 163219	163318	gap of 100 bp
* 163319	167426	contig of 4108 bp in length
* 167427	167526	gap of 100 bp
* 167527	172410	contig of 4884 bp in length
* 172411	172510	gap of 100 bp
* 172511	176689	contig of 4179 bp in length
* 176690	176789	gap of 100 bp
* 176790	180030	contig of 3241 bp in length
* 180031	180130	gap of 100 bp
* 180131	185024	contig of 4894 bp in length
* 185025	185124	gap of 100 bp
* 185125	188834	contig of 3710 bp in length
* 188835	188934	gap of 100 bp
* 188935	191247	contig of 2313 bp in length
* 191248	191347	gap of 100 bp
* 191348	194377	contig of 3030 bp in length
* 194378	194477	gap of 100 bp
* 194478	196487	contig of 2010 bp in length
* 196488	196587	gap of 100 bp
* 196588	199459	contig of 2872 bp in length
* 199460	199559	gap of 100 bp
* 199560	201207	contig of 1648 bp in length
* 201208	201307	gap of 100 bp
* 201308	203675	contig of 2368 bp in length
* 203676	203775	gap of 100 bp
* 203776	205300	contig of 1525 bp in length
* 205301	205400	gap of 100 bp
* 205401	207521	contig of 2121 bp in length

```
           *   207522   207621: gap of 100 bp
           *   207622   210083: contig of 2462 bp in length
           *   210084   210183: gap of 100 bp
           *   210184   211973: contig of 1790 bp in length.
FEATURES             Location/Qualifiers
     source          1. .211973
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q24"
                     /clone="RP11-820M9"
     misc_feature    1. .30856
                     /note="assembly_fragment"
     misc_feature    30957. .46182
                     /note="assembly_fragment clone_end:SP6 vector_side:right"
     misc_feature    46283. .60789
                     /note="assembly_fragment"
     misc_feature    60890. .73659
                     /note="assembly_fragment"
     misc_feature    73760. .83723
                     /note="assembly_fragment"
     misc_feature    83824. .93706
                     /note="assembly_fragment"
     misc_feature    93807. .103722
                     /note="assembly_fragment"
     misc_feature    103823. .112270
                     /note="assembly_fragment"
     misc_feature    112371. .119605
                     /note="assembly_fragment"
     misc_feature    119706. .124160
                     /note="assembly_fragment"
     misc_feature    124261. .129556
                     /note="assembly_fragment"
     misc_feature    129657. .134737
                     /note="assembly_fragment"
     misc_feature    134838. .139920
                     /note="assembly_fragment"
     misc_feature    140021. .145577
                     /note="assembly_fragment"
     misc_feature    145678. .149898
                     /note="assembly_fragment"
     misc_feature    149999. .154667

Alignment Scores:
Pred. No.:              3.78e-51        Length:         211973
Score:                  546.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-10-679-865A-2 (1-108) x AP001163 (1-211973)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db      50092 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCT 50151

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50152 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 50211

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50212 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 50271

Qy         61 LysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50272 AAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAAT 50331

Qy         81 ValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnValThrLeuLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50332 GTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTA 50391

Qy        101 CysAsnSerSerAlaGluAlaLeu 108
              |||||||||||||||||||||||
Db      50392 TGCAATTCCTCTGCTGAGGCCCTG 50415

RESULT 7
AC009544
LOCUS       AC009544               212357 bp    DNA     linear   HTG 22-APR-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-85D24 map 11, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC009544
VERSION     AC009544.4  GI:7637232
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 212357)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-85D24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212357)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 22, 2000 this sequence version replaced gi:6454062.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1649
                Center clone name: 85_D_24
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 3% of reads
                Chemistry: Dye-terminator Big Dye; 97% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 166296 bases at least Q40
                Consensus quality: 188982 bases at least Q30
                Consensus quality: 201949 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 211057; sum-of-contigs
                Quality coverage: 4.9 in Q20 bases; agarose-fp
                Quality coverage: 4.1 in Q20 base.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1174: contig of 1174 bp in length
```

```
            *     1175     1274: gap of 100 bp
            *     1275     2941: contig of 1667 bp in length
            *     2942     3041: gap of 100 bp
            *     3042     5780: contig of 2739 bp in length
            *     5781     5880: gap of 100 bp
            *     5881    12941: contig of 7061 bp in length
            *    12942    13041: gap of 100 bp
            *    13042    23383: contig of 10342 bp in length
            *    23384    23483: gap of 100 bp
            *    23484    32867: contig of 9384 bp in length
            *    32868    32967: gap of 100 bp
            *    32968    43838: contig of 10871 bp in length
            *    43839    43938: gap of 100 bp
            *    43939    55437: contig of 11499 bp in length
            *    55438    55537: gap of 100 bp
            *    55538    71158: contig of 15621 bp in length
            *    71159    71258: gap of 100 bp
            *    71259    87581: contig of 16323 bp in length
            *    87582    87681: gap of 100 bp
            *    87682   105529: contig of 17848 bp in length
            *   105530   105629: gap of 100 bp
            *   105630   120782: contig of 15153 bp in length
            *   120783   120882: gap of 100 bp
            *   120883   161108: contig of 40226 bp in length
            *   161109   161208: gap of 100 bp
            *   161209   212357: contig of 51149 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212357
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-85D24"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1174
                     /note="assembly_fragment"
     misc_feature    1275. .2941
                     /note="assembly_fragment"
     misc_feature    3042. .5780
                     /note="assembly_fragment"
     misc_feature    5881. .12941
                     /note="assembly_fragment"
     misc_feature    13042. .23383
                     /note="assembly_fragment"
     misc_feature    23484. .32867
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    32968. .43838
                     /note="assembly_fragment"
     misc_feature    43939. .55437
                     /note="assembly_fragment"
     misc_feature    55538. .71158
                     /note="assembly_fragment"
     misc_feature    71259. .87581
                     /note="assembly_fragment"
     misc_feature    87682. .105529
                     /note="assembly_fragment"
     misc_feature    105630. .120782
                     /note="assembly_fragment"
     misc_feature    120883. .161108
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    161209. .212357
                     /note="assembly_fragment"
ORIGIN

Alignment Scores:
Pred. No.:              3.79e-51        Length:         212357
Score:                  546.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
DB:                     2               Gaps:           0

US-10-679-865A-2 (1-108) x AC009544 (1-212357)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db      63790 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCT 63849

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63850 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 63909

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63910 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 63969

Qy         61 LysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63970 AAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAAT 64029

Qy         81 ValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnValThrLeuLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64030 GTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTA 64089

Qy        101 CysAsnSerSerAlaGluAlaLeu 108
              |||||||||||||||||||||||
Db      64090 TGCAATTCCTCTGCTGAGGCCCTG 64113

RESULT 8
AP001359/c
LOCUS       AP001359               221535 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q clone:RP11-820L6, complete
            sequences.
ACCESSION   AP001359
VERSION     AP001359.4  GI:14189751
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 221535)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On May 22, 2001 this sequence version replaced gi:9857628.
FEATURES             Location/Qualifiers
     source          1. .221535
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-820L6"
ORIGIN

Alignment Scores:
Pred. No.:              3.96e-51        Length:         221535
Score:                  546.00          Matches:        107
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  99.07%          Mismatches:     0
Query Match:            99.82%          Indels:         0
```

```
DB:                    9              Gaps:          0

US-10-679-865A-2 (1-108) x AP001359 (1-221535)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db      83880 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCT 83821

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83820 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 83761

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83760 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 83701

Qy         61 LysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83700 AAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAAT 83641

Qy         81 ValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnValThrLeuLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83640 GTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTA 83581

Qy        101 CysAsnSerSerAlaGluAlaLeu 108
              |||||||||||||||||||||||||
Db      83580 TGCAATTCCTCTGCTGAGGCCCTG 83557

RESULT 9
CQ738864
LOCUS       CQ738864               1572 bp    DNA     linear   PAT 03-FEB-2004
DEFINITION  Sequence 24798 from Patent WO02068579.
ACCESSION   CQ738864
VERSION     CQ738864.1  GI:42340434
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 24798 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .1572
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              3.23e-30       Length:         1572
Score:                  346.00         Matches:        72
Percent Similarity:     70.64%         Conservative:   5
Best Local Similarity:  66.06%         Mismatches:     6
Query Match:            63.25%         Indels:         26
DB:                     6              Gaps:           1

US-10-679-865A-2 (1-108) x CQ738864 (1-1572)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        220 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCT 279

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 339

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 399

Qy         61 LysGluThrValLeuSerLeu---------------------------------------- 67
              |||||||||||||||||||||
Db        400 AAGGAAACAGTGCTTTCTCTTAAAAGTGTGTTGGAAAACAGTGGAAGGACTGTTTGCCAA 459

Qy         68 -------------------------------------LysArgTyrAsnLeuGlySer 74
                                                   ||||||      |||||||||
Db        460 ATTGTCTTAGCGGTTCTCAAAGATGGCTGGCCAAATCATAAGAGAATACGCTTAGGTTCG 519

Qy         75 SerAlaMetLysArgAsnValProGly 83
                 :::   :::   :::::::   |||
Db        520 TGTAGTTCTGAGTCCAGCATTGTGGGT 546

RESULT 10
AC016859
LOCUS       AC016859               79544 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-13K20, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016859
VERSION     AC016859.2  GI:7144864
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 79544)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-13K20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 79544)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 3, 2000 this sequence version replaced gi:6539341.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3346
                Center clone name: 13_K_20
            --------------.
            * NOTE: This record contains 89 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      786: contig of 786 bp in length
            *      787      886: gap of 100 bp
```

*	887	1746:	contig of 860 bp in length
*	1747	1846:	gap of 100 bp
*	1847	2645:	contig of 799 bp in length
*	2646	2745:	gap of 100 bp
*	2746	3553:	contig of 808 bp in length
*	3554	3653:	gap of 100 bp
*	3654	4428:	contig of 775 bp in length
*	4429	4528:	gap of 100 bp
*	4529	5294:	contig of 766 bp in length
*	5295	5394:	gap of 100 bp
*	5395	6181:	contig of 787 bp in length
*	6182	6281:	gap of 100 bp
*	6282	7080:	contig of 799 bp in length
*	7081	7180:	gap of 100 bp
*	7181	7981:	contig of 801 bp in length
*	7982	8081:	gap of 100 bp
*	8082	8878:	contig of 797 bp in length
*	8879	8978:	gap of 100 bp
*	8979	9786:	contig of 808 bp in length
*	9787	9886:	gap of 100 bp
*	9887	10655:	contig of 769 bp in length
*	10656	10755:	gap of 100 bp
*	10756	11550:	contig of 795 bp in length
*	11551	11650:	gap of 100 bp
*	11651	12538:	contig of 888 bp in length
*	12539	12638:	gap of 100 bp
*	12639	13419:	contig of 781 bp in length
*	13420	13519:	gap of 100 bp
*	13520	14281:	contig of 762 bp in length
*	14282	14381:	gap of 100 bp
*	14382	15167:	contig of 786 bp in length
*	15168	15267:	gap of 100 bp
*	15268	16046:	contig of 779 bp in length
*	16047	16146:	gap of 100 bp
*	16147	16925:	contig of 779 bp in length
*	16926	17025:	gap of 100 bp
*	17026	17819:	contig of 794 bp in length
*	17820	17919:	gap of 100 bp
*	17920	18727:	contig of 808 bp in length
*	18728	18827:	gap of 100 bp
*	18828	19617:	contig of 790 bp in length
*	19618	19717:	gap of 100 bp
*	19718	20524:	contig of 807 bp in length
*	20525	20624:	gap of 100 bp
*	20625	21430:	contig of 806 bp in length
*	21431	21530:	gap of 100 bp
*	21531	22349:	contig of 819 bp in length
*	22350	22449:	gap of 100 bp
*	22450	23255:	contig of 806 bp in length
*	23256	23355:	gap of 100 bp
*	23356	24136:	contig of 781 bp in length
*	24137	24236:	gap of 100 bp
*	24237	25011:	contig of 775 bp in length
*	25012	25111:	gap of 100 bp
*	25112	25878:	contig of 767 bp in length
*	25879	25978:	gap of 100 bp
*	25979	26778:	contig of 800 bp in length
*	26779	26878:	gap of 100 bp
*	26879	27684:	contig of 806 bp in length
*	27685	27784:	gap of 100 bp
*	27785	28574:	contig of 790 bp in length
*	28575	28674:	gap of 100 bp
*	28675	29458:	contig of 784 bp in length
*	29459	29558:	gap of 100 bp
*	29559	30346:	contig of 788 bp in length
*	30347	30446:	gap of 100 bp
*	30447	31256:	contig of 810 bp in length
*	31257	31356:	gap of 100 bp
*	31357	32166:	contig of 810 bp in length
*	32167	32266:	gap of 100 bp
*	32267	33059:	contig of 793 bp in length
*	33060	33159:	gap of 100 bp
*	33160	33960:	contig of 801 bp in length
*	33961	34060:	gap of 100 bp
*	34061	34832:	contig of 772 bp in length
*	34833	34932:	gap of 100 bp
*	34933	35713:	contig of 781 bp in length
*	35714	35813:	gap of 100 bp
*	35814	36596:	contig of 783 bp in length
*	36597	36696:	gap of 100 bp
*	36697	37495:	contig of 799 bp in length
*	37496	37595:	gap of 100 bp
*	37596	38395:	contig of 800 bp in length
*	38396	38495:	gap of 100 bp
*	38496	39318:	contig of 823 bp in length
*	39319	39418:	gap of 100 bp
*	39419	40238:	contig of 820 bp in length
*	40239	40338:	gap of 100 bp
*	40339	41151:	contig of 813 bp in length
*	41152	41251:	gap of 100 bp
*	41252	42052:	contig of 801 bp in length
*	42053	42152:	gap of 100 bp
*	42153	42963:	contig of 811 bp in length
*	42964	43063:	gap of 100 bp
*	43064	43851:	contig of 788 bp in length
*	43852	43951:	gap of 100 bp
*	43952	44733:	contig of 782 bp in length
*	44734	44833:	gap of 100 bp
*	44834	45612:	contig of 779 bp in length
*	45613	45712:	gap of 100 bp
*	45713	46503:	contig of 791 bp in length
*	46504	46603:	gap of 100 bp
*	46604	47390:	contig of 787 bp in length
*	47391	47490:	gap of 100 bp
*	47491	48278:	contig of 788 bp in length
*	48279	48378:	gap of 100 bp
*	48379	49162:	contig of 784 bp in length
*	49163	49262:	gap of 100 bp
*	49263	50061:	contig of 799 bp in length
*	50062	50161:	gap of 100 bp
*	50162	50964:	contig of 803 bp in length
*	50965	51064:	gap of 100 bp
*	51065	51872:	contig of 808 bp in length
*	51873	51972:	gap of 100 bp
*	51973	52777:	contig of 805 bp in length
*	52778	52877:	gap of 100 bp
*	52878	53655:	contig of 778 bp in length
*	53656	53755:	gap of 100 bp
*	53756	54509:	contig of 754 bp in length
*	54510	54609:	gap of 100 bp
*	54610	55402:	contig of 793 bp in length
*	55403	55502:	gap of 100 bp
*	55503	56277:	contig of 775 bp in length
*	56278	56377:	gap of 100 bp
*	56378	57161:	contig of 784 bp in length
*	57162	57261:	gap of 100 bp
*	57262	58041:	contig of 780 bp in length
*	58042	58141:	gap of 100 bp
*	58142	58934:	contig of 793 bp in length
*	58935	59034:	gap of 100 bp
*	59035	59831:	contig of 797 bp in length
*	59832	59931:	gap of 100 bp
*	59932	60737:	contig of 806 bp in length
*	60738	60837:	gap of 100 bp
*	60838	61647:	contig of 810 bp in length
*	61648	61747:	gap of 100 bp
*	61748	62558:	contig of 811 bp in length
*	62559	62658:	gap of 100 bp
*	62659	63442:	contig of 784 bp in length
*	63443	63542:	gap of 100 bp
*	63543	64317:	contig of 775 bp in length
*	64318	64417:	gap of 100 bp
*	64418	65189:	contig of 772 bp in length

Alignment Scores:
Pred. No.: 1.66e-11 Length: 79544

```
Score:                  201.00          Matches:        38
Percent Similarity:     95.12%          Conservative:   1
Best Local Similarity:  92.68%          Mismatches:     2
Query Match:            36.75%          Indels:         0
DB:                     2               Gaps:           0

US-10-679-865A-2 (1-108) x AC016859 (1-79544)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              |||||||||||||||||   |||||||||||:::|||||||||||||||||||||||||
Db      39196 ATGGTGACTTTGTTGCCTTTTGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCT 39255

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              |||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db      39256 GAGTGTGTGCTCTACACACGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 39315

Qy         41 Ala 41
              |||
Db      39316 GCA 39318
```

```
RESULT 11
AE017319
LOCUS       AE017319               300349 bp    DNA     linear   BCT 27-APR-2004
DEFINITION  Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough, section
            11 of 12 of the complete genome.
ACCESSION   AE017319 AE017285
VERSION     AE017319.1  GI:46450738
KEYWORDS    .
SOURCE      Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough
  ORGANISM  Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough
            Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
            Desulfovibrionaceae; Desulfovibrio.
REFERENCE   1  (bases 1 to 300349)
  AUTHORS   Heidelberg,J.F., Seshadri,R., Haveman,S.A., Hemme,C.L.,
            Paulsen,I.T., Kolonay,J.F., Eisen,J.A., Ward,N., Methe,B.,
            Brinkac,L.M., Daugherty,S.C., DeBoy,R.T., Dodson,R.J., Durkin,A.S.,
            Madupu,R., Nelson,W.C., Sullivan,S.A., Fouts,D.E., Haft,D.H.,
            Selengut,J., Peterson,J.D., Davidsen,T.M., Zafar,N., Zhou,L.,
            Radune,D., Dimitrov,G., Hance,M., Tran,K., Khouri,H.M., Gill,J.,
            Utterback,T.R., Feldblyum,T.V., Wall,J.D., Voordouw,G. and
            Fraser,C.M.
  TITLE     The genome sequence of the anaerobic, sulfate-reducing bacterium
            Desulfovibrio vulgaris Hildenborough
  JOURNAL   Nat. Biotechnol. (2004) In press
   PUBMED   15077118
REFERENCE   2  (bases 1 to 300349)
  AUTHORS   Heidelberg,J.F., Seshadri,R., Haveman,S.A., Hemme,C.L.,
            Paulsen,I.T., Kolonay,J.F., Eisen,J.A., Ward,N., Methe,B.,
            Brinkac,L.M., Daugherty,S.C., DeBoy,R.T., Dodson,R.J., Durkin,A.S.,
            Madupu,R., Nelson,W.C., Sullivan,S.A., Fouts,D.E., Haft,D.H.,
            Selengut,J., Peterson,J.D., Davidsen,T.M., Zafar,N., Zhou,L.,
            Radune,D., Dimitrov,G., Hance,M., Tran,K., Khouri,H.M., Gill,J.,
            Utterback,T.R., Feldblyum,T.V., Wall,J.D., Voordouw,G. and
            Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .300349
                     /organism="Desulfovibrio vulgaris subsp. vulgaris str.
                     Hildenborough"
                     /mol_type="genomic DNA"
                     /strain="Hildenborough"
                     /sub_species="vulgaris"
                     /db_xref="taxon:882"
     gene            88. .1095
                     /locus_tag="DVU2913"
     CDS             88. .1095
                     /locus_tag="DVU2913"
                     /codon_start=1
                     /transl_table=11
                     /product="lipoprotein, putative"
                     /protein_id="AAS97385.1"
                     /db_xref="GI:46450739"
                     /translation="MIDHRPTSWSRLCTALTSGASCTIGGQAVMEGVMMRNGDRLAIA
                     MRRPDGEIVAESRPWFSLTRSAWLKRPFVRGFPILVETLVNGIKALNRSAEQAVEGET
                     GEELKPWQLAATLAVSVAMALGLFVVVPHLFSIGMKLLGLGGDVEGLSFHMWDGFFKF
                     AVFIGYIVAISFVPDIRRVFQYHGAEHKVIRAFEAAGDVTPDTAMHHSRLHPRCGTTF
                     LLFVLSISIILHAVLVPLLMLVWTPSGAFAKHGVTVLFKLLLMIPISALAYETIRYAA
                     RLGEGLTGTLLRAPGMLLQKLTTGEPDREQLEVAIVALREALGDDAPQSVRAPAYRHL
                     E"
     gene            1102. .2175
                     /gene="prfA"
                     /locus_tag="DVU2914"
     CDS             1102. .2175
                     /gene="prfA"
                     /locus_tag="DVU2914"
                     /note="identified by similarity to SP:P13654; match to
                     protein family HMM PF00472; match to protein family HMM
                     PF03462; match to protein family HMM TIGR00019"
                     /codon_start=1
                     /transl_table=11
                     /product="peptide chain release factor 1"
                     /protein_id="AAS97386.1"
                     /db_xref="GI:46450740"
                     /translation="MFAKLENLELKFEDLEQQLSSAEVFNDQDRYRKLTKAHADLKQV
                     VDAFRRYKEMKQNLADNKELLGDSDHEIRAMAHEEIKAIEAALPDIEQELKILLLPRD
                     PMDDKNILLEIRAGTGGEEASLFAADLFRMYTRYAEIMGWKVEVLSASDSDTGGYKEI
                     IALIAGDKVYSRLKYESGTHRVQRVPATEAQGRIHTSAATVAVMPEAEEVDVDIRPDD
                     LRIDVYRASGAGGQHVNKTESAVRITHLPTGIVVACQDEKSQHKNKAKAMKVLISRVL
                     QAEQERAHSVIADARRALVGSGDRSERIRTYNYPQSRITDHRINLTLYSLDKVMEGEL
                     APLVDALVTHAQTEALKAQADAS"
     gene            2486. .2650
                     /locus_tag="DVU2915"
     CDS             2486. .2650
                     /locus_tag="DVU2915"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAS97387.1"
                     /db_xref="GI:46450741"
                     /translation="MRGARIGCNARRGARGRAMTAYPADADAAVPPLGHMSPPVPLCS
                     RRRHAARRST"
     gene            2875. .3762
                     /gene="hemK"
                     /locus_tag="DVU2916"
     CDS             2875. .3762
                     /gene="hemK"
                     /locus_tag="DVU2916"
                     /note="identified by similarity to SP:P37186; match to
                     protein family HMM TIGR00536"
                     /codon_start=1
                     /transl_table=11
                     /product="hemK protein"
                     /protein_id="AAS97388.1"
                     /db_xref="GI:46450742"
                     /translation="MNPASTSHATLQGLLRHARQRLEAAEVDAPRLSAEIILCHALSL
                     RRIDIMLTPDRIVEEADCILFSELVARRATGEPLAYIVGEKEFFGRDFAVNPSTLIPR
                     PETEHLIETALESLRSGPARFVDAGTGSGCIAVTLCAERADLSGLALDMSAPALATAS
                     HNARRHGVAQRLAFVRGDFTTSLLRSGSLDLYASNPPYISEAEYTGLSREVRDFEPRS
                     ALVPGDTGLEHAAAIIAEATRVLRPHGILLMEFGCMQGADMASLFTPYSTLWEMVEVR
                     RDLAGLDRFIFARRNLLQP"
     gene            3970. .4896
                     /gene="lpxC"
                     /locus_tag="DVU2917"
     CDS             3970. .4896
                     /gene="lpxC"
                     /locus_tag="DVU2917"
                     /EC_number="3.5.1.-"
                     /note="identified by similarity to SP:P07652; match to
                     protein family HMM PF03331; match to protein family HMM
                     TIGR00325"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="UDP-3-0-acyl N-acetylglucosamine deacetylase"
                     /protein_id="AAS97389.1"
                     /db_xref="GI:46450743"
                     /translation="MNQTTIKKSIACSGVGLHSGKTVRMVLHPAAEDTGIVFDIHTAQ
                     GVRRIAPEPQVVIATGLATTLGMDGASVATVEHLLAAIRGLEIDNITVEIEGGEVPIM
                     DGSAASFVMLLRNAGIRRQTSARKVFRIARPVHYERDGKSIRALPYDGFRVEYRIEFP
                     HPLIGRQTLSIDITPESFGEIAKARTFGFLREVEYLHSKGLALGGSLDNAIVLDDYSV
                     LNPDGLRSPDEFVRHKVLDFVGDMAMMGVPLQGHFIVECSGHALNNGFLRMLEENASL
                     YLEAVELPVAEQHPAALRPAARVATEGQPAIA"
gene                 5029. .5265
                     /locus_tag="DVU2918"
CDS                  5029. .5265
                     /locus_tag="DVU2918"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAS97390.1"
                     /db_xref="GI:46450744"
                     /translation="MYDNWYGVQRRVPWLHITMAKTRDPIRVAATHHACHIAHIGRIR
                     HNPSEGPAHPLPAGADGVRHLLLRACRNAEVGYD"
gene                 complement(5228. .5767)
                     /locus_tag="DVU2919"
CDS                  complement(5228. .5767)
                     /locus_tag="DVU2919"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAS97391.1"
                     /db_xref="GI:46450745"
                     /translation="MLLYEGSWLGKHKNAFFRICFQVFFIYSVCKANFFSTIKVFTCA
                     LGTVSVGARLPRRGRGSCLLYCCVPLASPTLTVRIPSVLPAMVLVRAGACACALCPHW
                     ACFGGMSGYGVPDLCPEAGGGALGAFVQRQCPSVAFEGSRKWPSVLTYGLVSCALCVV
                     LLHGFAVFNHIPLPRCGKP"
gene                 5778. .5853
                     /locus_tag="DVU_tRNA-Thr-3"
tRNA                 5778. .5853
                     /locus_tag="DVU_tRNA-Thr-3"
                     /product="tRNA-Thr"
gene                 5860. .5945
                     /locus_tag="DVU_tRNA-Tyr-1"
tRNA                 5860. .5945
                     /locus_tag="DVU_tRNA-Tyr-1"
                     /product="tRNA-Tyr"
gene                 6003. .6079
                     /locus_tag="DVU_tRNA-Gly-5"
tRNA                 6003. .6079
                     /locus_tag="DVU_tRNA-Gly-5"
                     /product="tRNA-Gly"
gene                 6116. .6191
                     /locus_tag="DVU_tRNA-Thr-4"
tRNA                 6116. .6191
                     /locus_tag="DVU_tRNA-Thr-4"
                     /product="tRNA-Thr"
gene                 6244. .7437
                     /gene="tuf"
                     /locus_tag="DVU2920"
CDS                  6244. .7437
                     /gene="tuf"
                     /locus_tag="DVU2920"
                     /note="identified by similarity to SP:O50306; similarity
                     to SP:P02990; match to protein family HMM PF00009; match
                     to protein family HMM PF03143; match to protein family HMM
                     PF03144; match to protein family HMM TIGR00231; match to
                     protein family HMM TIGR00485"
                     /codon_start=1
                     /transl_table=11
                     /product="translation elongation factor Tu"
                     /protein_id="AAS97392.1"
                     /db_xref="GI:46450746"
                     /translation="MGKEKFERKKPHVNIGTIGHIDHGKTTLTAAITKTAGLLGQGKF
                     IAYDEIDKAPEEKERGITIATAHVEYETATRHYAHVDCPGHADYIKNMITGAAQMDGA
                     IIVVAATDGPMPQTREHILLARQVGVPYIVVFLNKCDMVDDEELLELVELEVRELLTS

Alignment Scores:
Pred. No.:              1.44e+03        Length:         300349
Score:                  86.50           Matches:        27
Percent Similarity:     43.33%          Conservative:   12
Best Local Similarity:  30.00%          Mismatches:     26
Query Match:            15.81%          Indels:         25
DB:                     1               Gaps:           4

US-10-679-865A-2 (1-108) x AE017319 (1-300349)

Qy         27 GlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArgLeuPro 46
              ||||||          |||      :::|||       |||            |||   |||
Db      62344 GGATGGGCGACCAGCGCAGACTCGACCTCGGCCGTGCCC---------ATGCGGTGTCCC 62394

Qy         47 LeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuProLysGluThrValLeuSer 66
                 |||   |||:::   |||::::::   ||||||:::   |||         :::|||
Db      62395 ---GAGACGTTGATGACGTCGTCGAGGCGGCCCATGATCCAGAAATAGCCGTCCGTATCG 62451

Qy         67 LeuLys----------------------------------------------------Arg 69
              ::::::                                                    |||
Db      62452 GTGCGTGCGCCATCGCCGGACTCGTACATGCCGGGAAAGCGTTCGAAGTATGTCGAGCGG 62511

Qy         70 TyrAsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgPro 89
              |||       |||          |||   ||||||:::|||||||||   |||:::
Db      62512 TACCGTTCGGGCGAGCCGAAGATGCCGCGCAACATGCCCGGCCACGGCTTGCGG------ 62565

Qy         90 SerTyrLeuThrArgIleGlnValThrLeu 99
                 :::||||||       :::|||
Db      62566 ------ATGACAAGATGCCCGCCTTCGTTG 62589

RESULT 12
MMU278829/c
LOCUS       MMU278829               1384 bp    mRNA    linear   ROD 14-SEP-2001
DEFINITION  Mus musculus mRNA for putative pyroglutamyl-peptidase I (PGPEP1
            gene).
ACCESSION   AJ278829
VERSION     AJ278829.1  GI:9843780
KEYWORDS    PGPEP1 gene; pyroglutamyl-peptidase I.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Dando,P.M., Fortunato,M., Rawlings,N.D. and Barrett,A.J.
  TITLE     Cloning, sequencing and expression of mouse pyroglutamyl-peptidase
            I
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1384)
  AUTHORS   Dando,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-2000) Dando P.M., MRC Molecular Enzymology
            Laboratory, The Babraham Institute, Cambridge, CB2 4AT, UNITED
            KINGDOM
COMMENT     Related sequence: AJ278828.
FEATURES             Location/Qualifiers
     source          1. .1384
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="I.M.A.G.E:2123102"
     gene            1. .1384
                     /gene="PGPEP1"
     CDS             28. .657
                     /gene="PGPEP1"
                     /EC_number="3.4.19.3"
                     /codon_start=1
                     /product="putative pyroglutamyl-peptidase I"
                     /protein_id="CAC03615.1"
                     /db_xref="GI:9843781"
```

```
                     /db_xref="GOA:Q9ESW8"
                     /db_xref="UniProt/Swiss-Prot:Q9ESW8"
                     /translation="MEQPRKAVVVTGFGPFGEHTVNASWIAVQELEKLGLGDSVDLHV
                     YEIPVEYQTVQRLIPALWEKHSPQLVVHVGVSGMATTVTLEKCGHNKGYKGLDNCRFC
                     PGSQCCVEDGPESIDSIIDMDAVCKRVTTLGLDVSVTISQDAGRYLCDFTYYTSLYQG
                     RGRSAFVHVPPLGKPYNADQLGRALRAIIEEMLGVLEQAEGDISCCRQL"
ORIGIN

Alignment Scores:
Pred. No.:                6.18          Length:        1384
Score:                    85.00         Matches:       24
Percent Similarity:       49.33%        Conservative:  13
Best Local Similarity:    32.00%        Mismatches:    34
Query Match:              15.54%        Indels:        4
DB:                       10            Gaps:          2

US-10-679-865A-2 (1-108) x MMU278829 (1-1384)

Qy         34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
              ||||||          |||||||||   |||       ||||||   |||:::
Db       1299 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCCATAGCAGCCTGC 1240

Qy         52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                    :::||||||   :::      ::::::::::::::
Db       1239 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 1180

Qy         71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                   :::|||:::   |||           |||:::|||   ||||||
Db       1179 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 1120

Qy         91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
              ||||||          |||||||||   |||:::|||:::
Db       1119 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 1075

RESULT 13
BC051938/c
LOCUS       BC051938               2161 bp    mRNA    linear   ROD 31-AUG-2004
DEFINITION  Mus musculus pyroglutamyl-peptidase I, mRNA (cDNA clone MGC:62180
            IMAGE:5717365), complete cds.
ACCESSION   BC051938
VERSION     BC051938.1  GI:30704646
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2161)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2161)
  AUTHORS   Director MGC Project.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library Preparation: M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 114 Row: k Column: 2
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 12963582.
FEATURES             Location/Qualifiers
     source          1. .2161
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="MGC:62180 IMAGE:5717365"
                     /tissue_type="Brain, mouse 12.5 dpc"
                     /clone_lib="NIH_BMAP_FD0"
                     /lab_host="DH10B"
                     /note="Vector: pYX-ASC"
     gene            1. .2161
                     /gene="Pgpep1"
                     /note="synonyms: PGP, Pcp, PGPI, PGP-I"
                     /db_xref="LocusID:66522"
                     /db_xref="MGI:1913772"
     CDS             25. .654
                     /gene="Pgpep1"
                     /codon_start=1
                     /product="putative pyroglutamyl-peptidase I"
                     /protein_id="AAH51938.1"
                     /db_xref="GI:30704647"
                     /db_xref="LocusID:66522"
                     /db_xref="MGI:1913772"
                     /translation="MEQPRKAVVVTGFGPFGEHTVNASWIAVQELEKLGLGDSVDLHV
                     YEIPVEYQTVQRLIPALWEKHSPQLVVHVGVSGMATTVTLEKCGHNKGYKGLDNCRFC
                     PGSQCCVEDGPESIDSIIDMDAVCKRVTTLGLDVSVTISQDAGRYLCDFTYYTSLYQG
                     RGRSAFVHVPPLGKPYNADQLGRALRAIIEEMLGVLEQAEGDISCCRQL"
ORIGIN

Alignment Scores:
Pred. No.:                10            Length:        2161
Score:                    85.00         Matches:       24
Percent Similarity:       49.33%        Conservative:  13
Best Local Similarity:    32.00%        Mismatches:    34
Query Match:              15.54%        Indels:        4
DB:                       10            Gaps:          2

US-10-679-865A-2 (1-108) x BC051938 (1-2161)

Qy         34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
              ||||||          |||||||||   |||       ||||||   |||:::
Db       1296 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCCATAGCAGCCTGC 1237

Qy         52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                    :::||||||   :::      ::::::::::::::
Db       1236 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 1177

Qy         71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                   :::|||:::   |||           |||:::|||   ||||||
Db       1176 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 1117
```

```
Qy          91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
               ||||||           |||||||||   |||:::|||:::
Db        1116 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 1072

RESULT 14
BC043041/c
LOCUS       BC043041               2170 bp    mRNA    linear   ROD 25-JUN-2004
DEFINITION  Mus musculus RIKEN cDNA 2810003H13 gene, mRNA (cDNA clone
            IMAGE:5701330), partial cds.
ACCESSION   BC043041
VERSION     BC043041.1  GI:28175475
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2170)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2170)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library Preparation: M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
            Thomas L. Casavant.
            Web site: http://genome.uiowa.edu
            Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
            Bonaldo,M.F., Akabogu,I.,  Bair,T., Bair,J., Crouch,K., Davis,A.,
            Fishler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
            Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J.,
            Casavant,T., Soares,M.B.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series:  Plate:  Row:  Column: 0
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 12963582.
FEATURES             Location/Qualifiers
     source          1. .2170
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5701330"
                     /tissue_type="Brain, mouse 15.5 dpc"
                     /clone_lib="NIH_BMAP_EV0"
                     /lab_host="DH10B"
                     /note="Vector: pYX-ASC"
ORIGIN

Alignment Scores:
Pred. No.:                10.1          Length:          2170
Score:                    85.00         Matches:         24
Percent Similarity:       49.33%        Conservative:    13
Best Local Similarity:    32.00%        Mismatches:      34
Query Match:              15.54%        Indels:          4
DB:                       10            Gaps:            2

US-10-679-865A-2 (1-108) x BC043041 (1-2170)

Qy          34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
               ||||||          |||||||||   |||       ||||||   |||:::
Db        1277 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCCATAGCAGCCTGC 1218

Qy          52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                     :::||||||   :::      ::::::::::::::
Db        1217 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 1158

Qy          71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                     :::|||:::   |||              |||:::|||   ||||||
Db        1157 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 1098

Qy          91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
               ||||||           |||||||||   |||:::|||:::
Db        1097 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 1053

RESULT 15
AC074154
LOCUS       AC074154               185630 bp    DNA     linear   HTG 02-SEP-2000
DEFINITION  Mus musculus clone RP23-179D8, WORKING DRAFT SEQUENCE, 16 ordered
            pieces.
ACCESSION   AC074154
VERSION     AC074154.2  GI:9964823
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 185630)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185630)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 2, 2000 this sequence version replaced gi:9211264.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1805688
            Center clone name: RPCI-23_179D8
            -------
            Summary Statistics
            Consensus quality: 165532 bases at least Q40
            Consensus quality: 179722 bases at least Q30
            Consensus quality: 182368 bases at least Q20
            Estimated insert size: 154340; agarose-fp estimation
            Estimated insert size: 184930; sum-of-contigs estimation
            Quality coverage:  7.26 in Q20 bases; agarose-fp estimation
            Quality coverage:  6.06 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. Gaps between the contigs
```

```
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1     7707: contig of 7707 bp in length
            *    7708     7807: gap of unknown length
            *    7808    17643: contig of 9836 bp in length
            *   17644    17743: gap of unknown length
            *   17744    22597: contig of 4854 bp in length
            *   22598    22697: gap of unknown length
            *   22698    31773: contig of 9076 bp in length
            *   31774    31873: gap of unknown length
            *   31874    51472: contig of 19599 bp in length
            *   51473    51572: gap of unknown length
            *   51573    53264: contig of 1692 bp in length
            *   53265    53364: gap of unknown length
            *   53365    55679: contig of 2315 bp in length
            *   55680    55779: gap of unknown length
            *   55780    60648: contig of 4869 bp in length
            *   60649    60748: gap of unknown length
            *   60749    90923: contig of 30175 bp in length
            *   90924    91023: gap of unknown length
            *   91024    92934: contig of 1911 bp in length
            *   92935    93034: gap of unknown length
            *   93035   111899: contig of 18865 bp in length
            *  111900   111999: gap of unknown length
            *  112000   118843: contig of 6844 bp in length
            *  118844   118943: gap of unknown length
            *  118944   134350: contig of 15407 bp in length
            *  134351   134450: gap of unknown length
            *  134451   136179: contig of 1729 bp in length
            *  136180   136279: gap of unknown length
            *  136280   175147: contig of 38868 bp in length
            *  175148   175247: gap of unknown length
            *  175248   185630: contig of 10383 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185630
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-179D8"
                     /clone_lib="RPCI mouse BAC library 23"
ORIGIN

Alignment Scores:
Pred. No.:              1.28e+03        Length:         185630
Score:                  85.00           Matches:        24
Percent Similarity:     49.33%          Conservative:   13
Best Local Similarity:  32.00%          Mismatches:     34
Query Match:            15.54%          Indels:         4
DB:                     2               Gaps:           2

US-10-679-865A-2 (1-108) x AC074154 (1-185630)

Qy          34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
               ||||||         |||||||||   |||      ||||||   |||:::
Db      182557 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCCATAGCAGCCTGC 182616

Qy          52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                  :::||||||   :::      ::::::::::::::
Db      182617 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 182676

Qy          71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                  :::|||:::   |||            |||:::|||   ||||||
Db      182677 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 182736

Qy          91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
               ||||||            |||||||||   |||:::|||:::
Db      182737 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 182781
```

Search completed: August 31, 2005, 23:02:03
Job time : 7926 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2005, 23:07:16 ; Search time 4122 Seconds
(without alignments)
10133.041 Million cell updates/sec

Title: US-10-679-865A-1
Perfect score: 862
Sequence: 1 atgatgttgtgtttggagaa.....taattaaaacaaatcttgca 862

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		862	100.0	862	9	AF303179	AF303179 Homo sapi
2	c	860.4	99.8	102807	2	AP000674	AP000674 Homo sapi
3		860.4	99.8	119419	2	AP000666	AP000666 Homo sapi
4	c	860.4	99.8	164784	2	AP001384	AP001384 Homo sapi
5	c	860.4	99.8	190360	9	AP001924	AP001924 Homo sapi
6		860.4	99.8	211973	2	AP001163	AP001163 Homo sapi
7		860.4	99.8	212357	2	AC009544	AC009544 Homo sapi
8	c	860.4	99.8	221535	9	AP001359	AP001359 Homo sapi
9		409.6	47.5	79544	2	AC016859	AC016859 Homo sapi
10		216.8	25.2	1572	6	CQ738864	CQ738864 Sequence
11		49.4	5.7	77526	10	BX005244	BX005244 Mouse DNA
12		48	5.6	232128	2	AC114656	AC114656 Mus muscu
13	c	46.2	5.4	207004	5	AL954695	AL954695 Zebrafish
14	c	45.8	5.3	229612	2	AC095534	AC095534 Rattus no
15		45.8	5.3	238644	2	AC115381	AC115381 Rattus no
16		45	5.2	153929	2	AC026023	AC026023 Homo sapi
17	c	44	5.1	42724	3	CEF53A2	Z81546 Caenorhabdi
18		43.8	5.1	414	3	AY498556	AY498556 Lipolexis
19		43.4	5.0	2000	6	AX655393	AX655393 Sequence
20	c	43.4	5.0	154482	10	AC126430	AC126430 Mus muscu
21		43.4	5.0	232571	10	AC116557	AC116557 Mus muscu
22	c	43.2	5.0	202553	2	AC128197	AC128197 Rattus no
23		43.2	5.0	251930	10	AL772293	AL772293 Mouse DNA
24	c	42.8	5.0	6103	6	AX346592	AX346592 Sequence
25	c	42.8	5.0	8873	6	AX251820	AX251820 Sequence
26	c	42.8	5.0	8873	6	AX344206	AX344206 Sequence
27	c	42.8	5.0	8873	6	AX348605	AX348605 Sequence
28	c	42.6	4.9	2000	6	AX655393	AX655393 Sequence
29	c	42.6	4.9	163710	2	AC112834	AC112834 Rattus no
30	c	42.6	4.9	240033	2	AC133684	AC133684 Rattus no
31		42.4	4.9	82139	3	AC115684	AC115684 Dictyoste
32	c	42	4.9	247875	2	AC137235	AC137235 Rattus no
33		42	4.9	249541	2	AC106560	AC106560 Rattus no
34	c	42	4.9	250480	2	AC115644	AC115644 Rattus no
35	c	41.8	4.8	1141	6	AX083744	AX083744 Sequence
36		41.8	4.8	7300	3	CEY48A6A	AL023843 Caenorhab
37	c	41.8	4.8	35435	2	CEH06F14	Z92792 Caenorhabdi
38		41.8	4.8	144207	2	AC120530	AC120530 Oryza sat
39	c	41.8	4.8	197516	8	AC146948	AC146948 Oryza sat
40	c	41.6	4.8	477	6	AR545611	AR545611 Sequence
41	c	41.6	4.8	4401	8	MEBGLA	X94986 M.esculenta
42		41.6	4.8	97905	8	AC006192	AC006192 Arabidops
43		41.6	4.8	117927	2	AC145109	AC145109 Medicago
44		41.6	4.8	129894	2	AC108181	AC108181 Felis cat
45		41.6	4.8	169598	2	AC129245	AC129245 Rattus no

ALIGNMENTS

RESULT 1
AF303179
LOCUS AF303179 862 bp mRNA linear PRI 14-JUN-2004
DEFINITION Homo sapiens BRCC2 mRNA, complete cds.
ACCESSION AF303179
VERSION AF303179.1 GI:23393067
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 862)
AUTHORS Broustas,C.G., Gokhale,P.C., Rahman,A., Dritschilo,A., Ahmad,I. and Kasid,U.
TITLE BRCC2, a Novel BH3-like-domain-containing protein, induces apoptosis in a caspase-dependent manner
JOURNAL J. Biol. Chem. 279 (25), 26780-26788 (2004)
PUBMED 15069058
REFERENCE 2 (bases 1 to 862)
AUTHORS Gokhale,P.C., Kumar,D., Monia,B.P., Rahman,A., Dritschilo,A. and Kasid,U.
TITLE Novel Transcript in Human Breast Cancer Cells
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 862)
AUTHORS Gokhale,P.C., Rahman,A., Dritschilo,A. and Kasid,U.
TITLE Direct Submission
JOURNAL Submitted (05-SEP-2000) Radiation Medicine, Georgetown University, 3970 Reservoir Rd, NW, Washington, DC 20007, USA
FEATURES Location/Qualifiers
source 1. .862
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="11"
/tissue_type="breast cancer"
CDS 294. .620
/codon_start=1
/product="BRCC2"
/protein_id="AAN31127.1"
/db_xref="GI:23393068"
/translation="MVTLLPIEGQEVHFFEILESECVLYTGWIERASGSSIYPEAKAR
LPLEALLGSNKEPMLPKETVLSLKRYNLGSSAMKRNVPGHVLQRPSYLTRIQVTLLCN

SSAEAL"
ORIGIN

Query Match 100.0%; Score 862; DB 9; Length 862;
Best Local Similarity 100.0%; Pred. No. 1.4e-209;
Matches 862; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540

Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720

Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||
Db        841 TTTAATTAAAACAAATCTTGCA 862
```

RESULT 2
AP000674/c
LOCUS AP000674 102807 bp DNA linear HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 11 clone CMB9-8G24 map 11q24, WORKING DRAFT SEQUENCE, 17 unordered pieces.

ACCESSION AP000674
VERSION AP000674.2 GI:8118863
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 102807)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 102,807 genomic DNA of 11q24
JOURNAL Published Only in DataBase (1999)
REFERENCE 2 (bases 1 to 102807)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1999) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT On May 31, 2000 this sequence version replaced gi:6997548.
---------- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
---------- Project Information
Center project name: HumDraft11
Center clone name: CMB9-8G24
---------- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 90471 bases at least Q40
Consensus quality: 96773 bases at least Q30
Consensus quality: 99873 bases at least Q20
Insert size: 101207; sum-of-contigs
Quality coverage: 4.21x in Q20 bases; sum-of-contigs
----------.
NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1	23670	contig of	23670	bp in length
23771	39763	contig of	15993	bp in length
39864	50918	contig of	11055	bp in length
51019	61370	contig of	10352	bp in length
61471	70516	contig of	9046	bp in length
70617	75763	contig of	5147	bp in length
75864	81132	contig of	5269	bp in length
81233	84386	contig of	3154	bp in length
84487	87524	contig of	3038	bp in length
87625	89403	contig of	1779	bp in length
89504	91991	contig of	2488	bp in length
92092	94299	contig of	2208	bp in length
94400	96329	contig of	1930	bp in length
96430	97921	contig of	1492	bp in length
98022	100412	contig of	2391	bp in length
100513	101594	contig of	1082	bp in length
101695	102807	contig of	1113	bp in length

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```
          * be preserved.
          *        1    23670: contig of 23670 bp in length
          *    23671    23770: gap of 100 bp
          *    23771    39763: contig of 15993 bp in length
          *    39764    39863: gap of 100 bp
          *    39864    50918: contig of 11055 bp in length
          *    50919    51018: gap of 100 bp
          *    51019    61370: contig of 10352 bp in length
          *    61371    61470: gap of 100 bp
          *    61471    70516: contig of 9046 bp in length
          *    70517    70616: gap of 100 bp
          *    70617    75763: contig of 5147 bp in length
          *    75764    75863: gap of 100 bp
          *    75864    81132: contig of 5269 bp in length
          *    81133    81232: gap of 100 bp
          *    81233    84386: contig of 3154 bp in length
          *    84387    84486: gap of 100 bp
          *    84487    87524: contig of 3038 bp in length
          *    87525    87624: gap of 100 bp
          *    87625    89403: contig of 1779 bp in length
          *    89404    89503: gap of 100 bp
          *    89504    91991: contig of 2488 bp in length
          *    91992    92091: gap of 100 bp
          *    92092    94299: contig of 2208 bp in length
          *    94300    94399: gap of 100 bp
          *    94400    96329: contig of 1930 bp in length
          *    96330    96429: gap of 100 bp
          *    96430    97921: contig of 1492 bp in length
          *    97922    98021: gap of 100 bp
          *    98022   100412: contig of 2391 bp in length
          *   100413   100512: gap of 100 bp
          *   100513   101594: contig of 1082 bp in length
          *   101595   101694: gap of 100 bp
          *   101695   102807: contig of 1113 bp in length.
FEATURES             Location/Qualifiers
     source          1. .102807
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q24"
                     /clone="CMB9-8G24"
     misc_feature    1. .23670
                     /note="assembly_fragment"
     misc_feature    23771. .39763
                     /note="assembly_fragment"
     misc_feature    39864. .50918
                     /note="assembly_fragment"
     misc_feature    51019. .61370
                     /note="assembly_fragment"
     misc_feature    61471. .70516
                     /note="assembly_fragment"
     misc_feature    70617. .75763
                     /note="assembly_fragment"
     misc_feature    75864. .81132
                     /note="assembly_fragment"
     misc_feature    81233. .84386
                     /note="assembly_fragment"
     misc_feature    84487. .87524
                     /note="assembly_fragment"
     misc_feature    87625. .89403
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    89504. .91991
                     /note="assembly_fragment"
     misc_feature    92092. .94299
                     /note="assembly_fragment"
     misc_feature    94400. .96329
                     /note="assembly_fragment"
     misc_feature    96430. .97921
                     /note="assembly_fragment"
     misc_feature    98022. .100412
                     /note="assembly_fragment"
     misc_feature    100513. .101594
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    101695. .102807
                     /note="assembly_fragment"
ORIGIN

  Query Match             99.8%;  Score 860.4;  DB 2;  Length 102807;
  Best Local Similarity   99.9%;  Pred. No. 3.8e-209;
  Matches  861;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31921 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 31862

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31861 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 31802

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31801 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 31742

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31741 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 31682

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31681 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 31622

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      31621 CTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCTGAGTGTG 31562

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31561 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 31502

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31501 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 31442

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31441 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 31382

Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31381 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 31322

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31321 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 31262

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31261 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 31202

Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31201 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 31142

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31141 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 31082

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||
Db      31081 TTTAATTAAAACAAATCTTGCA 31060


RESULT 3
AP000666
LOCUS       AP000666               119419 bp    DNA     linear   HTG 30-MAY-2000
```

```
DEFINITION  Homo sapiens chromosome 11 clone CMB9-53P8 map 11q23, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AP000666
VERSION     AP000666.2  GI:8118855
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 119419)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 119,419 genomic DNA of 11q23
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 119419)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997540.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: CMB9-53P8
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 109205 bases at least Q40
                Consensus quality: 113734 bases at least Q30
                Consensus quality: 116089 bases at least Q20
                Insert size: 117519; sum-of-contigs
                Quality coverage: 4.59x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            20 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     22810 contig of  22810 bp in length
              22911     38820 contig of  15910 bp in length
              38921     52522 contig of  13602 bp in length
              52623     62588 contig of   9966 bp in length
              62689     72547 contig of   9859 bp in length
              72648     78455 contig of   5808 bp in length
              78556     83981 contig of   5426 bp in length
              84082     89379 contig of   5298 bp in length
              89480     93708 contig of   4229 bp in length
              93809     96503 contig of   2695 bp in length
              96604    100510 contig of   3907 bp in length
             100611    104144 contig of   3534 bp in length
             104245    107321 contig of   3077 bp in length
             107422    109944 contig of   2523 bp in length
             110045    111930 contig of   1886 bp in length
             112031    113584 contig of   1554 bp in length
             113685    115373 contig of   1689 bp in length
             115474    117116 contig of   1643 bp in length
             117217    118797 contig of   1581 bp in length
             118898    119419 contig of    522 bp in length
            Sequence updated (26-May-2000).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    22810: contig of 22810 bp in length
            *   22811    22910: gap of 100 bp
            *   22911    38820: contig of 15910 bp in length
            *   38821    38920: gap of 100 bp
            *   38921    52522: contig of 13602 bp in length
            *   52523    52622: gap of 100 bp
            *   52623    62588: contig of 9966 bp in length
            *   62589    62688: gap of 100 bp
            *   62689    72547: contig of 9859 bp in length
            *   72548    72647: gap of 100 bp
            *   72648    78455: contig of 5808 bp in length
            *   78456    78555: gap of 100 bp
            *   78556    83981: contig of 5426 bp in length
            *   83982    84081: gap of 100 bp
            *   84082    89379: contig of 5298 bp in length
            *   89380    89479: gap of 100 bp
            *   89480    93708: contig of 4229 bp in length
            *   93709    93808: gap of 100 bp
            *   93809    96503: contig of 2695 bp in length
            *   96504    96603: gap of 100 bp
            *   96604   100510: contig of 3907 bp in length
            *  100511   100610: gap of 100 bp
            *  100611   104144: contig of 3534 bp in length
            *  104145   104244: gap of 100 bp
            *  104245   107321: contig of 3077 bp in length
            *  107322   107421: gap of 100 bp
            *  107422   109944: contig of 2523 bp in length
            *  109945   110044: gap of 100 bp
            *  110045   111930: contig of 1886 bp in length
            *  111931   112030: gap of 100 bp
            *  112031   113584: contig of 1554 bp in length
            *  113585   113684: gap of 100 bp
            *  113685   115373: contig of 1689 bp in length
            *  115374   115473: gap of 100 bp
            *  115474   117116: contig of 1643 bp in length
            *  117117   117216: gap of 100 bp
            *  117217   118797: contig of 1581 bp in length
            *  118798   118897: gap of 100 bp
            *  118898   119419: contig of 522 bp in length.
FEATURES             Location/Qualifiers
     source          1. .119419
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q23"
                     /clone="CMB9-53P8"
     misc_feature    1. .22810
                     /note="assembly_fragment"
     misc_feature    22911. .38820
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    38921. .52522
                     /note="assembly_fragment"
     misc_feature    52623. .62588
                     /note="assembly_fragment"
     misc_feature    62689. .72547
                     /note="assembly_fragment"
     misc_feature    72648. .78455
                     /note="assembly_fragment"
     misc_feature    78556. .83981
                     /note="assembly_fragment"
     misc_feature    84082. .89379
                     /note="assembly_fragment"
     misc_feature    89480. .93708
                     /note="assembly_fragment"
     misc_feature    93809. .96503
                     /note="assembly_fragment"
```

```
     misc_feature    96604. .100510
                     /note="assembly_fragment"
     misc_feature    100611. .104144
                     /note="assembly_fragment"
     misc_feature    104245. .107321
                     /note="assembly_fragment"
     misc_feature    107422. .109944
                     /note="assembly_fragment"
     misc_feature    110045. .111930
                     /note="assembly_fragment"
     misc_feature    112031. .113584
                     /note="assembly_fragment"
     misc_feature    113685. .115373
                     /note="assembly_fragment"
     misc_feature    115474. .117116
                     /note="assembly_fragment"
     misc_feature    117217. .118797
                     /note="assembly_fragment"
     misc_feature    118898. .119419
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
ORIGIN

  Query Match             99.8%;  Score 860.4;  DB 2;  Length 119419;
  Best Local Similarity   99.9%;  Pred. No. 3.9e-209;
  Matches  861;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74360 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 74419

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74420 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 74479

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74480 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 74539

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74540 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 74599

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74600 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 74659

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      74660 CTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCTGAGTGTG 74719

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74720 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 74779

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74780 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 74839

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74840 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 74899

Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74900 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 74959

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74960 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 75019

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75020 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 75079

Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75080 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 75139

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75140 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 75199

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||
Db      75200 TTTAATTAAAACAAATCTTGCA 75221


RESULT 4
AP001384/c
LOCUS       AP001384              164784 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-382J20 map 11q24, WORKING
            DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION   AP001384
VERSION     AP001384.2  GI:8117296
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164784)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 164,784 genomic DNA of 11q24
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 164784)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7243295.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: RP11-382J20
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 147339 bases at least Q40
                Consensus quality: 156937 bases at least Q30
                Consensus quality: 160723 bases at least Q20
                Insert size: 162984; sum-of-contigs
                Quality coverage: 4.61x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            19 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     29112 contig of  29112 bp in length
              29213     60097 contig of  30885 bp in length
              60198     74677 contig of  14480 bp in length
              74778     90573 contig of  15796 bp in length
              90674     99271 contig of   8598 bp in length
              99372    108087 contig of   8716 bp in length
```

```
            108188    116782 contig of   8595 bp in length
            116883    124764 contig of   7882 bp in length
            124865    130409 contig of   5545 bp in length
            130510    137878 contig of   7369 bp in length
            137979    143242 contig of   5264 bp in length
            143343    148670 contig of   5328 bp in length
            148771    151308 contig of   2538 bp in length
            151409    154990 contig of   3582 bp in length
            155091    157252 contig of   2162 bp in length
            157353    159636 contig of   2284 bp in length
            159737    161763 contig of   2027 bp in length
            161864    163588 contig of   1725 bp in length
            163689    164784 contig of   1096 bp in length
        Sequence updated (26-May-2000).
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 19 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *        1    29112: contig of 29112 bp in length
        *    29113    29212: gap of 100 bp
        *    29213    60097: contig of 30885 bp in length
        *    60098    60197: gap of 100 bp
        *    60198    74677: contig of 14480 bp in length
        *    74678    74777: gap of 100 bp
        *    74778    90573: contig of 15796 bp in length
        *    90574    90673: gap of 100 bp
        *    90674    99271: contig of 8598 bp in length
        *    99272    99371: gap of 100 bp
        *    99372   108087: contig of 8716 bp in length
        *   108088   108187: gap of 100 bp
        *   108188   116782: contig of 8595 bp in length
        *   116783   116882: gap of 100 bp
        *   116883   124764: contig of 7882 bp in length
        *   124765   124864: gap of 100 bp
        *   124865   130409: contig of 5545 bp in length
        *   130410   130509: gap of 100 bp
        *   130510   137878: contig of 7369 bp in length
        *   137879   137978: gap of 100 bp
        *   137979   143242: contig of 5264 bp in length
        *   143243   143342: gap of 100 bp
        *   143343   148670: contig of 5328 bp in length
        *   148671   148770: gap of 100 bp
        *   148771   151308: contig of 2538 bp in length
        *   151309   151408: gap of 100 bp
        *   151409   154990: contig of 3582 bp in length
        *   154991   155090: gap of 100 bp
        *   155091   157252: contig of 2162 bp in length
        *   157253   157352: gap of 100 bp
        *   157353   159636: contig of 2284 bp in length
        *   159637   159736: gap of 100 bp
        *   159737   161763: contig of 2027 bp in length
        *   161764   161863: gap of 100 bp
        *   161864   163588: contig of 1725 bp in length
        *   163589   163688: gap of 100 bp
        *   163689   164784: contig of 1096 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164784
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q24"
                     /clone="RP11-382J20"
     misc_feature    1. .29112
                     /note="assembly_fragment"
     misc_feature    29213. .60097
                     /note="assembly_fragment"
     misc_feature    60198. .74677
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    74778. .90573
                     /note="assembly_fragment"
     misc_feature    90674. .99271
                     /note="assembly_fragment"
     misc_feature    99372. .108087
                     /note="assembly_fragment"
     misc_feature    108188. .116782
                     /note="assembly_fragment"
     misc_feature    116883. .124764
                     /note="assembly_fragment"
     misc_feature    124865. .130409
                     /note="assembly_fragment"
     misc_feature    130510. .137878
                     /note="assembly_fragment"
     misc_feature    137979. .143242
                     /note="assembly_fragment"
     misc_feature    143343. .148670
                     /note="assembly_fragment"
     misc_feature    148771. .151308
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    151409. .154990
                     /note="assembly_fragment"
     misc_feature    155091. .157252
                     /note="assembly_fragment"
     misc_feature    157353. .159636
                     /note="assembly_fragment"
     misc_feature    159737. .161763
                     /note="assembly_fragment"
     misc_feature    161864. .163588
                     /note="assembly_fragment"
     misc_feature    163689. .164784
                     /note="assembly_fragment"
ORIGIN

  Query Match             99.8%;  Score 860.4;  DB 2;  Length 164784;
  Best Local Similarity   99.9%;  Pred. No. 3.9e-209;
  Matches  861;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21044 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 20985

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20984 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 20925

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20924 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 20865

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20864 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 20805

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20804 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 20745

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      20744 CTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCTGAGTGTG 20685

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20684 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 20625

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20624 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 20565

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20564 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 20505
```

```
Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20504 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 20445

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20444 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 20385

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20384 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 20325

Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20324 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 20265

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20264 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 20205

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||||
Db      20204 TTTAATTAAAACAAATCTTGCA 20183
```

```
RESULT 5
AP001924/c
LOCUS       AP001924               190360 bp    DNA     linear   PRI 01-FEB-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11 clone:RP11-166D19, complete
            sequence.
ACCESSION   AP001924
VERSION     AP001924.5  GI:28189518
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 190360)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Jan 31, 2003 this sequence version replaced gi:19879815.
FEATURES             Location/Qualifiers
     source          1. .190360
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-166D19"
ORIGIN

  Query Match             99.8%;  Score 860.4;  DB 9;  Length 190360;
  Best Local Similarity   99.9%;  Pred. No. 3.9e-209;
  Matches  861;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165813 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 165754

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165753 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 165694

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165693 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 165634

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165633 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 165574

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165573 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 165514

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     165513 CTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCTGAGTGTG 165454

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165453 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 165394

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165393 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 165334

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165333 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 165274

Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165273 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 165214

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165213 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 165154

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165153 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 165094

Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165093 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 165034

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165033 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 164974

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||||
Db     164973 TTTAATTAAAACAAATCTTGCA 164952
```

```
RESULT 6
AP001163
LOCUS       AP001163               211973 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-820M9 map 11q24, WORKING
            DRAFT SEQUENCE, 34 unordered pieces.
ACCESSION   AP001163
VERSION     AP001163.2  GI:8118731
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 211973)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 211,973 genomic DNA of 11q24
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 211973)
```

```
AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
          Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE     Direct Submission
JOURNAL   Submitted (10-FEB-2000) Masahira Hattori, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
          Japan (E-mail:hattori@gsc.riken.go.jp,
          URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
          Fax:81-42-778-9924)
COMMENT   On May 31, 2000 this sequence version replaced gi:6997849.
          -------------- Genome Center
              Center: RIKEN Genomic Sciences Center(GSC)
              Center code: RIKEN
              Web site: http://hgp.gsc.riken.go.jp/
              Contact: hattori@gsc.riken.go.jp
          -------------- Project Information
              Center project name: HumDraft11
              Center clone name: RP11-820M9
          -------------- Summary Statistics
              Sequencing vector: PCR products; 100% of reads
              Chemistry: Dye-terminator ET-amersham; 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 185228 bases at least Q40
              Consensus quality: 199175 bases at least Q30
              Consensus quality: 205608 bases at least Q20
              Insert size: 208673; sum-of-contigs
              Quality coverage: 4.21x in Q20 bases; sum-of-contigs
          --------------.
          NOTE: This is a 'working draft' sequence. It currently consists of
          34 contigs. The true order of the pieces is not known and their
          order in this sequence record is arbitrary. Gaps between the
          contigs are represented as runs N, but the exact sizes of the gaps
          are unknown. This record will be updated with the finished sequence
          as soon as it is available and the accession number will be
          preserved
                1     30856 contig of  30856 bp in length
            30957     46182 contig of  15226 bp in length
            46283     60789 contig of  14507 bp in length
            60890     73659 contig of  12770 bp in length
            73760     83723 contig of   9964 bp in length
            83824     93706 contig of   9883 bp in length
            93807    103722 contig of   9916 bp in length
           103823    112270 contig of   8448 bp in length
           112371    119605 contig of   7235 bp in length
           119706    124160 contig of   4455 bp in length
           124261    129556 contig of   5296 bp in length
           129657    134737 contig of   5081 bp in length
           134838    139920 contig of   5083 bp in length
           140021    145577 contig of   5557 bp in length
           145678    149898 contig of   4221 bp in length
           149999    154667 contig of   4669 bp in length
           154768    158909 contig of   4142 bp in length
           159010    163218 contig of   4209 bp in length
           163319    167426 contig of   4108 bp in length
           167527    172410 contig of   4884 bp in length
           172511    176689 contig of   4179 bp in length
           176790    180030 contig of   3241 bp in length
           180131    185024 contig of   4894 bp in length
           185125    188834 contig of   3710 bp in length
           188935    191247 contig of   2313 bp in length
           191348    194377 contig of   3030 bp in length
           194478    196487 contig of   2010 bp in length
           196588    199459 contig of   2872 bp in length
           199560    201207 contig of   1648 bp in length
           201308    203675 contig of   2368 bp in length
           203776    205300 contig of   1525 bp in length
           205401    207521 contig of   2121 bp in length
           207622    210083 contig of   2462 bp in length
           210184    211973 contig of   1790 bp in length
          Sequence updated (26-May-2000).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 34 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    30856: contig of 30856 bp in length
          *   30857    30956: gap of 100 bp
          *   30957    46182: contig of 15226 bp in length
          *   46183    46282: gap of 100 bp
          *   46283    60789: contig of 14507 bp in length
          *   60790    60889: gap of 100 bp
          *   60890    73659: contig of 12770 bp in length
          *   73660    73759: gap of 100 bp
          *   73760    83723: contig of 9964 bp in length
          *   83724    83823: gap of 100 bp
          *   83824    93706: contig of 9883 bp in length
          *   93707    93806: gap of 100 bp
          *   93807   103722: contig of 9916 bp in length
          *  103723   103822: gap of 100 bp
          *  103823   112270: contig of 8448 bp in length
          *  112271   112370: gap of 100 bp
          *  112371   119605: contig of 7235 bp in length
          *  119606   119705: gap of 100 bp
          *  119706   124160: contig of 4455 bp in length
          *  124161   124260: gap of 100 bp
          *  124261   129556: contig of 5296 bp in length
          *  129557   129656: gap of 100 bp
          *  129657   134737: contig of 5081 bp in length
          *  134738   134837: gap of 100 bp
          *  134838   139920: contig of 5083 bp in length
          *  139921   140020: gap of 100 bp
          *  140021   145577: contig of 5557 bp in length
          *  145578   145677: gap of 100 bp
          *  145678   149898: contig of 4221 bp in length
          *  149899   149998: gap of 100 bp
          *  149999   154667: contig of 4669 bp in length
          *  154668   154767: gap of 100 bp
          *  154768   158909: contig of 4142 bp in length
          *  158910   159009: gap of 100 bp
          *  159010   163218: contig of 4209 bp in length
          *  163219   163318: gap of 100 bp
          *  163319   167426: contig of 4108 bp in length
          *  167427   167526: gap of 100 bp
          *  167527   172410: contig of 4884 bp in length
          *  172411   172510: gap of 100 bp
          *  172511   176689: contig of 4179 bp in length
          *  176690   176789: gap of 100 bp
          *  176790   180030: contig of 3241 bp in length
          *  180031   180130: gap of 100 bp
          *  180131   185024: contig of 4894 bp in length
          *  185025   185124: gap of 100 bp
          *  185125   188834: contig of 3710 bp in length
          *  188835   188934: gap of 100 bp
          *  188935   191247: contig of 2313 bp in length
          *  191248   191347: gap of 100 bp
          *  191348   194377: contig of 3030 bp in length
          *  194378   194477: gap of 100 bp
          *  194478   196487: contig of 2010 bp in length
          *  196488   196587: gap of 100 bp
          *  196588   199459: contig of 2872 bp in length
          *  199460   199559: gap of 100 bp
          *  199560   201207: contig of 1648 bp in length
          *  201208   201307: gap of 100 bp
          *  201308   203675: contig of 2368 bp in length
          *  203676   203775: gap of 100 bp
          *  203776   205300: contig of 1525 bp in length
          *  205301   205400: gap of 100 bp
          *  205401   207521: contig of 2121 bp in length
          *  207522   207621: gap of 100 bp
          *  207622   210083: contig of 2462 bp in length
          *  210084   210183: gap of 100 bp
          *  210184   211973: contig of 1790 bp in length.
FEATURES             Location/Qualifiers
```

```
     source          1. .211973
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q24"
                     /clone="RP11-820M9"
     misc_feature    1. .30856
                     /note="assembly_fragment"
     misc_feature    30957. .46182
                     /note="assembly_fragment clone_end:SP6 vector_side:right"
     misc_feature    46283. .60789
                     /note="assembly_fragment"
     misc_feature    60890. .73659
                     /note="assembly_fragment"
     misc_feature    73760. .83723
                     /note="assembly_fragment"
     misc_feature    83824. .93706
                     /note="assembly_fragment"
     misc_feature    93807. .103722
                     /note="assembly_fragment"
     misc_feature    103823. .112270
                     /note="assembly_fragment"
     misc_feature    112371. .119605
                     /note="assembly_fragment"
     misc_feature    119706. .124160
                     /note="assembly_fragment"
     misc_feature    124261. .129556
                     /note="assembly_fragment"
     misc_feature    129657. .134737
                     /note="assembly_fragment"
     misc_feature    134838. .139920
                     /note="assembly_fragment"
     misc_feature    140021. .145577
                     /note="assembly_fragment"
     misc_feature    145678. .149898
                     /note="assembly_fragment"
     misc_feature    149999. .154667

  Query Match             99.8%;  Score 860.4;  DB 2;  Length 211973;
  Best Local Similarity   99.9%;  Pred. No. 3.9e-209;
  Matches  861;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49799 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 49858

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49859 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 49918

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49919 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 49978

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49979 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 50038

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50039 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 50098

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      50099 CTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCTGAGTGTG 50158

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50159 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 50218

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      50219 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 50278

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50279 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 50338

Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50339 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 50398

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50399 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 50458

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50459 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 50518

Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50519 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 50578

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50579 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 50638

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||
Db      50639 TTTAATTAAAACAAATCTTGCA 50660
```

RESULT 7
AC009544

```
LOCUS       AC009544               212357 bp    DNA     linear   HTG 22-APR-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-85D24 map 11, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC009544
VERSION     AC009544.4  GI:7637232
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 212357)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-85D24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212357)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 22, 2000 this sequence version replaced gi:6454062.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
```

```
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
            Center project name: L1649
            Center clone name: 85_D_24
          -------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-primer-amersham; 3% of reads
            Chemistry: Dye-terminator Big Dye; 97% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 166296 bases at least Q40
            Consensus quality: 188982 bases at least Q30
            Consensus quality: 201949 bases at least Q20
            Insert size: 176000; agarose-fp
            Insert size: 211057; sum-of-contigs
            Quality coverage: 4.9 in Q20 bases; agarose-fp
            Quality coverage: 4.1 in Q20 base.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 14 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1174: contig of 1174 bp in length
          *     1175     1274: gap of 100 bp
          *     1275     2941: contig of 1667 bp in length
          *     2942     3041: gap of 100 bp
          *     3042     5780: contig of 2739 bp in length
          *     5781     5880: gap of 100 bp
          *     5881    12941: contig of 7061 bp in length
          *    12942    13041: gap of 100 bp
          *    13042    23383: contig of 10342 bp in length
          *    23384    23483: gap of 100 bp
          *    23484    32867: contig of 9384 bp in length
          *    32868    32967: gap of 100 bp
          *    32968    43838: contig of 10871 bp in length
          *    43839    43938: gap of 100 bp
          *    43939    55437: contig of 11499 bp in length
          *    55438    55537: gap of 100 bp
          *    55538    71158: contig of 15621 bp in length
          *    71159    71258: gap of 100 bp
          *    71259    87581: contig of 16323 bp in length
          *    87582    87681: gap of 100 bp
          *    87682   105529: contig of 17848 bp in length
          *   105530   105629: gap of 100 bp
          *   105630   120782: contig of 15153 bp in length
          *   120783   120882: gap of 100 bp
          *   120883   161108: contig of 40226 bp in length
          *   161109   161208: gap of 100 bp
          *   161209   212357: contig of 51149 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212357
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-85D24"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1174
                     /note="assembly_fragment"
     misc_feature    1275. .2941
                     /note="assembly_fragment"
     misc_feature    3042. .5780
                     /note="assembly_fragment"
     misc_feature    5881. .12941
                     /note="assembly_fragment"
     misc_feature    13042. .23383
                     /note="assembly_fragment"
     misc_feature    23484. .32867
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    32968. .43838
                     /note="assembly_fragment"
     misc_feature    43939. .55437
                     /note="assembly_fragment"
     misc_feature    55538. .71158
                     /note="assembly_fragment"
     misc_feature    71259. .87581
                     /note="assembly_fragment"
     misc_feature    87682. .105529
                     /note="assembly_fragment"
     misc_feature    105630. .120782
                     /note="assembly_fragment"
     misc_feature    120883. .161108
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    161209. .212357
                     /note="assembly_fragment"
ORIGIN

  Query Match             99.8%;  Score 860.4;  DB 2;  Length 212357;
  Best Local Similarity   99.9%;  Pred. No. 3.9e-209;
  Matches  861;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63497 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 63556

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63557 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 63616

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63617 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 63676

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63677 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 63736

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63737 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 63796

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      63797 CTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCTGAGTGTG 63856

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63857 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 63916

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63917 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 63976

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63977 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 64036

Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64037 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 64096

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64097 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 64156

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64157 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 64216
```

```
Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64217 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 64276

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64277 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 64336

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||||
Db      64337 TTTAATTAAAACAAATCTTGCA 64358
```

RESULT 8
AP001359/c
LOCUS AP001359 221535 bp DNA linear PRI 15-MAR-2003
DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-820L6, complete sequences.
ACCESSION AP001359
VERSION AP001359.4 GI:14189751
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 221535)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On May 22, 2001 this sequence version replaced gi:9857628.
FEATURES Location/Qualifiers
source 1. .221535
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-820L6"

ORIGIN

Query Match 99.8%; Score 860.4; DB 9; Length 221535;
Best Local Similarity 99.9%; Pred. No. 3.9e-209;
Matches 861; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84173 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 84114

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84113 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 84054

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84053 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 83994

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83993 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 83934

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83933 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 83874

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83873 CTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCTGAGTGTG 83814

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83813 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 83754

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83753 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 83694

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83693 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 83634

Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83633 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 83574

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83573 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 83514

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83513 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 83454

Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83453 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 83394

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83393 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 83334

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||||
Db      83333 TTTAATTAAAACAAATCTTGCA 83312
```

RESULT 9
AC016859
LOCUS AC016859 79544 bp DNA linear HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-13K20, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016859
VERSION AC016859.2 GI:7144864
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 79544)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-13K20
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 79544)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE Direct Submission
JOURNAL Submitted (08-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 3, 2000 this sequence version replaced gi:6539341.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L3346
Center clone name: 13_K_20
----------.
* NOTE: This record contains 89 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

*	1	786: contig of 786 bp in length
*	787	886: gap of 100 bp
*	887	1746: contig of 860 bp in length
*	1747	1846: gap of 100 bp
*	1847	2645: contig of 799 bp in length
*	2646	2745: gap of 100 bp
*	2746	3553: contig of 808 bp in length
*	3554	3653: gap of 100 bp
*	3654	4428: contig of 775 bp in length
*	4429	4528: gap of 100 bp
*	4529	5294: contig of 766 bp in length
*	5295	5394: gap of 100 bp
*	5395	6181: contig of 787 bp in length
*	6182	6281: gap of 100 bp
*	6282	7080: contig of 799 bp in length
*	7081	7180: gap of 100 bp
*	7181	7981: contig of 801 bp in length
*	7982	8081: gap of 100 bp
*	8082	8878: contig of 797 bp in length
*	8879	8978: gap of 100 bp
*	8979	9786: contig of 808 bp in length
*	9787	9886: gap of 100 bp
*	9887	10655: contig of 769 bp in length
*	10656	10755: gap of 100 bp
*	10756	11550: contig of 795 bp in length
*	11551	11650: gap of 100 bp
*	11651	12538: contig of 888 bp in length
*	12539	12638: gap of 100 bp
*	12639	13419: contig of 781 bp in length
*	13420	13519: gap of 100 bp
*	13520	14281: contig of 762 bp in length
*	14282	14381: gap of 100 bp
*	14382	15167: contig of 786 bp in length
*	15168	15267: gap of 100 bp
*	15268	16046: contig of 779 bp in length
*	16047	16146: gap of 100 bp
*	16147	16925: contig of 779 bp in length
*	16926	17025: gap of 100 bp
*	17026	17819: contig of 794 bp in length
*	17820	17919: gap of 100 bp
*	17920	18727: contig of 808 bp in length
*	18728	18827: gap of 100 bp
*	18828	19617: contig of 790 bp in length
*	19618	19717: gap of 100 bp
*	19718	20524: contig of 807 bp in length
*	20525	20624: gap of 100 bp
*	20625	21430: contig of 806 bp in length
*	21431	21530: gap of 100 bp
*	21531	22349: contig of 819 bp in length
*	22350	22449: gap of 100 bp
*	22450	23255: contig of 806 bp in length
*	23256	23355: gap of 100 bp
*	23356	24136: contig of 781 bp in length
*	24137	24236: gap of 100 bp
*	24237	25011: contig of 775 bp in length
*	25012	25111: gap of 100 bp
*	25112	25878: contig of 767 bp in length
*	25879	25978: gap of 100 bp
*	25979	26778: contig of 800 bp in length
*	26779	26878: gap of 100 bp
*	26879	27684: contig of 806 bp in length
*	27685	27784: gap of 100 bp
*	27785	28574: contig of 790 bp in length
*	28575	28674: gap of 100 bp
*	28675	29458: contig of 784 bp in length
*	29459	29558: gap of 100 bp
*	29559	30346: contig of 788 bp in length
*	30347	30446: gap of 100 bp
*	30447	31256: contig of 810 bp in length
*	31257	31356: gap of 100 bp
*	31357	32166: contig of 810 bp in length
*	32167	32266: gap of 100 bp
*	32267	33059: contig of 793 bp in length
*	33060	33159: gap of 100 bp
*	33160	33960: contig of 801 bp in length
*	33961	34060: gap of 100 bp
*	34061	34832: contig of 772 bp in length
*	34833	34932: gap of 100 bp
*	34933	35713: contig of 781 bp in length
*	35714	35813: gap of 100 bp
*	35814	36596: contig of 783 bp in length
*	36597	36696: gap of 100 bp
*	36697	37495: contig of 799 bp in length
*	37496	37595: gap of 100 bp
*	37596	38395: contig of 800 bp in length
*	38396	38495: gap of 100 bp
*	38496	39318: contig of 823 bp in length
*	39319	39418: gap of 100 bp
*	39419	40238: contig of 820 bp in length
*	40239	40338: gap of 100 bp
*	40339	41151: contig of 813 bp in length
*	41152	41251: gap of 100 bp
*	41252	42052: contig of 801 bp in length
*	42053	42152: gap of 100 bp
*	42153	42963: contig of 811 bp in length
*	42964	43063: gap of 100 bp
*	43064	43851: contig of 788 bp in length
*	43852	43951: gap of 100 bp
*	43952	44733: contig of 782 bp in length
*	44734	44833: gap of 100 bp
*	44834	45612: contig of 779 bp in length
*	45613	45712: gap of 100 bp
*	45713	46503: contig of 791 bp in length
*	46504	46603: gap of 100 bp
*	46604	47390: contig of 787 bp in length
*	47391	47490: gap of 100 bp
*	47491	48278: contig of 788 bp in length
*	48279	48378: gap of 100 bp
*	48379	49162: contig of 784 bp in length
*	49163	49262: gap of 100 bp
*	49263	50061: contig of 799 bp in length
*	50062	50161: gap of 100 bp
*	50162	50964: contig of 803 bp in length
*	50965	51064: gap of 100 bp
*	51065	51872: contig of 808 bp in length
*	51873	51972: gap of 100 bp
*	51973	52777: contig of 805 bp in length
*	52778	52877: gap of 100 bp
*	52878	53655: contig of 778 bp in length

```
            *    53656    53755: gap of 100 bp
            *    53756    54509: contig of 754 bp in length
            *    54510    54609: gap of 100 bp
            *    54610    55402: contig of 793 bp in length
            *    55403    55502: gap of 100 bp
            *    55503    56277: contig of 775 bp in length
            *    56278    56377: gap of 100 bp
            *    56378    57161: contig of 784 bp in length
            *    57162    57261: gap of 100 bp
            *    57262    58041: contig of 780 bp in length
            *    58042    58141: gap of 100 bp
            *    58142    58934: contig of 793 bp in length
            *    58935    59034: gap of 100 bp
            *    59035    59831: contig of 797 bp in length
            *    59832    59931: gap of 100 bp
            *    59932    60737: contig of 806 bp in length
            *    60738    60837: gap of 100 bp
            *    60838    61647: contig of 810 bp in length
            *    61648    61747: gap of 100 bp
            *    61748    62558: contig of 811 bp in length
            *    62559    62658: gap of 100 bp
            *    62659    63442: contig of 784 bp in length
            *    63443    63542: gap of 100 bp
            *    63543    64317: contig of 775 bp in length
            *    64318    64417: gap of 100 bp
            *    64418    65189: contig of 772 bp in length

  Query Match             47.5%;  Score 409.6;  DB 2;  Length 79544;
  Best Local Similarity   99.0%;  Pred. No. 7.4e-94;
  Matches  412;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38903 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 38962

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38963 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 39022

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39023 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 39082

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39083 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 39142

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      39143 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 39202

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              |||||||||| | |||||||||| ||||||||||||||||||||||||||||||||||||
Db      39203 CTTTGTTGCCTTTTGAGGGCCAGGAAATACATTTCTTTGAGATCCTAGAATCTGAGTGTG 39262

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCA 416
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db      39263 TGCTCTACACACGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCA 39318


RESULT 10
CQ738864
LOCUS       CQ738864               1572 bp    DNA     linear   PAT 03-FEB-2004
DEFINITION  Sequence 24798 from Patent WO02068579.
ACCESSION   CQ738864
VERSION     CQ738864.1  GI:42340434
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 24798 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .1572
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             25.2%;  Score 216.8;  DB 6;  Length 1572;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-44;
  Matches  218;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        280 AATAGCTTCAAATTATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTG 339
              || |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        206 AAGAGCTTCAAATTATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAATACATTTCTTTG 265

Qy        340 AGATCCTAGAATCTGAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTT 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 AGATCCTAGAATCTGAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTT 325

Qy        400 CCATTTATCCAGAGGCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAG 459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 CCATTTATCCAGAGGCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAG 385

Qy        460 AACCTATGTTGCCTAAGGAAACAGTGCTTTCTCTTAAAAG 499
              |||||||||||||||||||||||||||||||||||||||
Db        386 AACCTATGTTGCCTAAGGAAACAGTGCTTTCTCTTAAAAG 425


RESULT 11
BX005244
LOCUS       BX005244               77526 bp    DNA     linear   ROD 25-APR-2003
DEFINITION  Mouse DNA sequence from clone RP23-134F14 on chromosome X, complete
            sequence.
ACCESSION   BX005244
VERSION     BX005244.12  GI:30141724
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 77526)
  AUTHORS   Chapman,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 26, 2003 this sequence version replaced gi:29691854.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
```

```
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-134F14 is
            from the RPCI-23 Mouse BAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES              Location/Qualifiers
     source           1. .77526
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10090"
                      /chromosome="X"
                      /clone="RP23-134F14"
                      /clone_lib="RPCI-23"
ORIGIN

  Query Match              5.7%;  Score 49.4;  DB 10;  Length 77526;
  Best Local Similarity   54.0%;  Pred. No. 0.097;
  Matches  101;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        669 CCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAA 728
              || | | | ||   | | |||| ||||  | |  || || | || |||   |||    |
Db      73614 CCTTAAGTTAGAAAGTAAAACAGAGAAATACATATGTTTCCTATTAGAAAAAATGGAGAG 73673

Qy        729 CAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAG 788
              |   ||||||    | ||| |        |||   ||   || |   |||  ||  | ||
Db      73674 AATCATTTTCTGATAAATTTCAAGATGTTCTTAAACAGCATTTGGATAAAAAGTAAAAAA 73733

Qy        789 AAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTA 848
              |||| |||  |   |||| ||| ||||| | |||  ||||| ||   | |   |   |
Db      73734 TTACAAAAAACGTGGAAAAATGGAGAAACTTGATAGAAAAAAGAAACCTGGACAAATGGA 73793

Qy        849 AAACAAA 855
              ||||| |
Db      73794 AAACAGA 73800


RESULT 12
AC114656
LOCUS       AC114656               232128 bp    DNA     linear   HTG 12-MAR-2004
DEFINITION  Mus musculus chromosome 12 clone RP24-161A16 map 12, *** SEQUENCING
            IN PROGRESS ***, 6 unordered pieces.
ACCESSION   AC114656
VERSION     AC114656.5  GI:45384658
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 232128)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 12, clone RP24-161A16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 232128)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 232128)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2004 this sequence version replaced gi:30017903.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L24415
                Center clone name: 161_A_16
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    34530: contig of 34530 bp in length
            *    34531    34630: gap of 100 bp
            *    34631    82454: contig of 47824 bp in length
            *    82455    82554: gap of 100 bp
            *    82555   122018: contig of 39464 bp in length
            *   122019   122118: gap of 100 bp
            *   122119   166005: contig of 43887 bp in length
            *   166006   166105: gap of 100 bp
            *   166106   204014: contig of 37909 bp in length
            *   204015   204114: gap of 100 bp
            *   204115   232128: contig of 28014 bp in length.
FEATURES              Location/Qualifiers
     source           1. .232128
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10090"
```

```
                     /chromosome="12"
                     /map="12"
                     /clone="RP24-161A16"
                     /clone_lib="RPCI-24 Male Mouse BAC"
ORIGIN

  Query Match              5.6%;  Score 48;  DB 2;  Length 232128;
  Best Local Similarity   54.5%;  Pred. No. 0.23;
  Matches   96;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        684 CACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCC 743
              | |  || ||| | ||||| |||||  ||  ||||  ||||  |  |||  | | |
Db     169570 CTCTGCAGAGATGTAAACTGGATTTTTATTTGACCATATTCACACTAGTAATATTTGTAA 169629

Qy        744 ACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTT 803
              |||| ||   | ||| | |  || ||   ||| || ||||    |  |||| | || |
Db     169630 ACTTATATTTAAAACGTATTAAACTTATTTTCTAAATGTGAGTATCAACATTGAATTCGG 169689

Qy        804 TAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTT 859
              || |  |||  | |     | |    ||| |  |  ||| |   ||  | |||||
Db     169690 TATACTGGAAGAGCCAATAAATACTCAGTCACAGAACTTTTACCTAATTATATCTT 169745


RESULT 13
AL954695/c
LOCUS       AL954695              207004 bp    DNA     linear   VRT 24-OCT-2003
DEFINITION  Zebrafish DNA sequence from clone CH211-202N12 in linkage group 1,
            complete sequence.
ACCESSION   AL954695
VERSION     AL954695.14  GI:37699978
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 207004)
  AUTHORS   Garnett,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 16, 2003 this sequence version replaced gi:37496370.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU). For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml
            CH211-202N12 is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .207004
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-202N12"
                     /clone_lib="CHORI-211"
ORIGIN

  Query Match              5.4%;  Score 46.2;  DB 5;  Length 207004;
  Best Local Similarity   57.1%;  Pred. No. 0.65;
  Matches   84;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        715 GACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTT 774
              | |  |||||  |  || | |||  ||| |  || | ||| ||   ||||||  || ||
Db     116449 GGCTGAATTCAAATGAGCCATTTTAACCTAGATTAATCCAGAATAATTCCAAGATTAGTG 116390

Qy        775 CAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAA 834
                | |  |  |||  || ||| ||  | |    |   | || |||  || ||| ||||||
Db     116389 AGATTAATCTAGATTAAAAATGAATCTATGCCCACCTACAATCTGAATATTTATAAGTAA 116330

Qy        835 ATGGCATTTAATTAAAACAAATCTTGC 861
              | ||||| |  |  |  ||| | | |
Db     116329 TTAGCATTAACATCCAGAAAAACATCC 116303


RESULT 14
AC095534/c
LOCUS       AC095534              229612 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-8E9, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC095534
VERSION     AC095534.7  GI:30467405
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 229612)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
```

Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 229612)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 229612)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 9, 2003 this sequence version replaced gi:24940746.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GCQF
Center clone name: CH230-8E9
---------- Summary Statistics
Assembly program: Atlas;
Consensus quality: 209009 bases at least Q40
Consensus quality: 212738 bases at least Q30
Consensus quality: 215945 bases at least Q20
Estimated insert size: 226482; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 229612: contig of 229612 bp in length.

FEATURES Location/Qualifiers
source 1. .229612
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-8E9"
misc_feature 73945. .76308
/note="wgs_contig"
misc_feature 150591. .152157
/note="wgs_contig"
ORIGIN

Query Match 5.3%; Score 45.8; DB 2; Length 229612;
Best Local Similarity 51.2%; Pred. No. 0.82;
Matches 107; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

```
Qy      613 CCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACAGATTTCGCCCAT 672
            || || |||   |||||||| | ||| ||| |  |      | |      ||| |    |
Db    35276 CCTTGCAAAGAAAGTGAGTTTGTACATATGTACAAGATACATGGCTATCTTTTTGTTACT 35217

Qy      673 TATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGT 732
             | |        |  |    |   |   |||||||| || || ||  || | |||   ||
Db    35216 AAATGCCAAAATAAGAATATGCTAGTTTCTTGATTTAAAAAACACAAAAATATTATATGT 35157

Qy      733 CTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAAC 792
            |  |      | ||||      |||  | ||| ||||| ||   | |      ||||| ||
Db    35156 CAATATATACAACTTCCCTTAATACTTATTCTAAATTGGGGCAAGAAATAATAGAGAGAC 35097

Qy      793 AATAAAATTTTTAAAAAGGATAAACTGGC 821
            | ||| || |||||| |    |||   ||
Db    35096 ACTAATATATTTAAATACATGAAAACAGC 35068
```

RESULT 15
AC115381
LOCUS AC115381 238644 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-224M2, WORKING DRAFT SEQUENCE, 3 unordered pieces.
ACCESSION AC115381
VERSION AC115381.4 GI:30578732
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 238644)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,

```
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238644)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 238644)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:22857380.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GLZC
                Center clone name: CH230-224M2
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 221476 bases at least Q40
                Consensus quality: 224408 bases at least Q30
                Consensus quality: 226367 bases at least Q20
                Estimated insert size: 232116; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   235088: contig of 235088 bp in length
            *  235089   235188: gap of unknown length
            *  235189   236253: contig of 1065 bp in length
            *  236254   236353: gap of unknown length
            *  236354   238644: contig of 2291 bp in length.
FEATURES            Location/Qualifiers
     source         1. .238644
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-224M2"
     misc_feature   1. .1288
                    /note="wgs_end_extension
                    clone_end:Sp6"
     misc_feature   1339. .3095
                    /note="wgs_end_extension
                    clone_end:Sp6"
     misc_feature   4833. .6088
                    /note="wgs_end_extension
                    clone_end:Sp6"
     misc_feature   complement(7373. .8267)
                    /note="clone_boundary
                    clone_end:Sp6
                    site:EcoRI
                    end_sequence:BZ097629"
     misc_feature   complement(233774. .234625)
                    /note="clone_boundary
                    clone_end:T7
                    site:EcoRI
                    end_sequence:BZ097627"
ORIGIN

  Query Match             5.3%;  Score 45.8;  DB 2;  Length 238644;
  Best Local Similarity   51.2%;  Pred. No. 0.82;
  Matches  107;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy        613 CCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACAGATTTCGCCCAT 672
              || || |||   |||||||| | ||| ||| | |  |       | |      ||| |    |
Db      27271 CCTTGCAAAGAAAGTGAGTTTGTACATATGTACAAGATACATGGCTATCTTTTTGTTACT 27330

Qy        673 TATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGT 732
                | |        | |    |   |   |||||||| || || ||  || | |||   ||
Db      27331 AAATGCCAAAATAAGAATATGCTAGTTTCTTGATTTAAAAAACACAAAAATATTATATGT 27390

Qy        733 CTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAAC 792
              | |     | ||||      |||  | ||| ||||| ||   | |      ||||| ||
Db      27391 CAATATATACAACTTCCCTTAATACTTATTCTAAATTGGGGCAAGAAATAATAGAGAGAC 27450

Qy        793 AATAAAATTTTTAAAAAGGATAAACTGGC 821
              | ||| || ||||||  |    |||   ||
Db      27451 ACTAATATATTTAAATACATGAAAACAGC 27479


Search completed: September  1, 2005, 01:29:02
Job time : 4131 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 31, 2005, 20:50:42 ; Search time 2657 Seconds
(without alignments)
1547.212 Million cell updates/sec

Title: US-10-679-865A-2
Perfect score: 547
Sequence: 1 MVTLLPIEGQEVHFFEILES.....PSYLTRIQVTLLCNSSAEAL 108

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US10679865/runat_30082005_150801_15917/app_query.fasta_1.263
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10679865_@CGN_1_1_3437_@runat_30082005_150801_15917 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	87	15.9	428	1	AA079394	AA079394 zm96a01.s
c	2	86.5	15.8	642	6	CF024497	CF024497 QBS5g11.x
c	3	86.5	15.8	686	7	CF919839	CF919839 EST4139 Z
c	4	86.5	15.8	721	8	CC332766	CC332766 OGOAU48TV
	5	86.5	15.8	797	8	CC149584	CC149584 ZMMBBb001
c	6	86.5	15.8	859	9	CC673519	CC673519 OGVEL95TV
c	7	86	15.7	530	2	BB757526	BB757526 BB757526
	8	85	15.5	313	7	CR758635	CR758635 CR758635
c	9	85	15.5	313	7	CR759545	CR759545 CR759545
c	10	85	15.5	382	6	BY585586	BY585586 BY585586
	11	85	15.5	391	2	BF455298	BF455298 UI-M-BZ1-
c	12	85	15.5	429	5	BY443486	BY443486 BY443486
	13	85	15.5	445	1	AA000493	AA000493 mg23b04.r
c	14	85	15.5	453	5	BM939154	BM939154 UI-M-BZ1-
	15	85	15.5	454	2	BE449045	BE449045 ut85g04.y
c	16	85	15.5	470	2	BB759932	BB759932 BB759932
	17	85	15.5	472	1	AI158291	AI158291 ud27a04.r
c	18	85	15.5	476	2	BB744816	BB744816 BB744816
c	19	85	15.5	491	7	W80322	W80322 me90e08.r1
	20	85	15.5	503	5	BX525163	BX525163 BX525163
	21	85	15.5	535	1	AA152869	AA152869 mq56d02.r
	22	85	15.5	544	2	BE626697	BE626697 uu10d06.y
c	23	85	15.5	547	7	W10599	W10599 ma44b12.r1
	24	85	15.5	552	1	AI314325	AI314325 uj29f06.x
	25	85	15.5	560	1	AI314860	AI314860 uj25g08.x
	26	85	15.5	567	1	AI314875	AI314875 uj26a06.x
	27	85	15.5	660	2	BF467987	BF467987 UI-M-CD0-
c	28	85	15.5	734	7	CK633761	CK633761 UI-M-HN0-
c	29	85	15.5	1366	3	AK012658	AK012658 Mus muscu
c	30	85	15.5	2099	3	AK049632	AK049632 Mus muscu
c	31	85	15.5	2151	3	AK047835	AK047835 Mus muscu
c	32	83	15.2	536	6	CD688193	CD688193 EST4715 h
c	33	83	15.2	920	6	CB954203	CB954203 AGENCOURT
	34	82.5	15.1	884	2	BE959046	BE959046 601644809
c	35	82	15.0	300	1	AV119510	AV119510 AV119510
c	36	82	15.0	372	6	BY585050	BY585050 BY585050
c	37	82	15.0	433	6	BY585415	BY585415 BY585415
	38	81.5	14.9	718	8	BH140220	BH140220 ZMMBBb000
c	39	81.5	14.9	849	1	AI529732	AI529732 ui81g05.y
c	40	81.5	14.9	865	5	BU909125	BU909125 AGENCOURT
c	41	81.5	14.9	885	9	CNS04JQ3	AL293844 Tetraodon
	42	81	14.8	497	1	AI881980	AI881980 ul56g08.x
c	43	81	14.8	642	5	BQ694025	BQ694025 1000168 H
	44	81	14.8	697	4	BG823779	BG823779 602726986
c	45	80.5	14.7	462	8	AQ516405	AQ516405 HS_5081_B

ALIGNMENTS

RESULT 1
AA079394
LOCUS AA079394 428 bp mRNA linear EST 23-DEC-1997
DEFINITION zm96a01.s1 Stratagene colon HT29 (#937221) Homo sapiens cDNA clone IMAGE:545736 3', mRNA sequence.
ACCESSION AA079394
VERSION AA079394.1 GI:1618288
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 428)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
PUBMED 8889549
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
WARNING: There is evidence that suggests that the 384-well parent plate of this clone contains both human and mouse derived clones. Thus, the origin of this clone is uncertain. This caution should be

kept in mind should you use this clone.

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1257 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 321.

```
FEATURES             Location/Qualifiers
     source          1. .428
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3923925"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:545736"
                     /dev_stage="HT-29 (HTB 38) adenocarcinoma"
                     /lab_host="SOLR (kanamycin resistant)"
                     /clone_lib="Stratagene colon HT29 (#937221)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally. Primer: Oligo dT. From
                     adenocarcinoma HT-29 (HTB 38), moderately differentiated
                     grade II. Average insert size: 1.5 kb; Uni-ZAP XR Vector;
                     ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
ORIGIN

Alignment Scores:
Pred. No.:              2.93          Length:        428
Score:                  87.00         Matches:       25
Percent Similarity:     49.33%        Conservative:  12
Best Local Similarity:  33.33%        Mismatches:    34
Query Match:            15.90%        Indels:        4
DB:                     1             Gaps:          2

US-10-679-865A-2 (1-108) x AA079394 (1-428)

Qy         34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
              ||||||       |||||||||   |||      ||||||   ||||||
Db         86 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCATTAGCAGCCTGC 145

Qy         52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                     :::||||||   :::      ::::::::::::::
Db        146 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 205

Qy         71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                    :::|||:::   |||           |||:::|||   ||||||
Db        206 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 265

Qy         91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
              ||||||          |||||||||   |||:::|||:::
Db        266 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 310
```

```
RESULT 2
CF024497/c
LOCUS       CF024497                642 bp    mRNA    linear   EST 17-JUL-2003
DEFINITION  QBS5g11.xg QBS Zea mays cDNA clone QBS5g11, mRNA sequence.
ACCESSION   CF024497
VERSION     CF024497.1  GI:32919685
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   Genoplante.
  TITLE     Genoplante, a major partnership french program in plant genomics
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Genoplante
            Genoplante
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplante' (http://www.genoplante.com
            and http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="F2"
                     /db_xref="taxon:4577"
                     /clone="QBS5g11"
                     /tissue_type="roots extremities"
                     /clone_lib="QBS"
ORIGIN

Alignment Scores:
Pred. No.:              6.02          Length:        642
Score:                  86.50         Matches:       28
Percent Similarity:     46.51%        Conservative:  12
Best Local Similarity:  32.56%        Mismatches:    45
Query Match:            15.81%        Indels:        1
DB:                     6             Gaps:          1

US-10-679-865A-2 (1-108) x CF024497 (1-642)

Qy         16 GluIleLeu---GluSerGluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGly 34
              |||:::|||      |||   |||                |||   |||   :::
Db        393 GAAGTCCTAGGACGCAGTTCTTGCAAAGGGTGTGGCCGCACCATAGCACGTAAGGCACAT 334

Qy         35 SerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsn 54
              |||   :::|||   :::             |||   :::      ||||||      :::
Db        333 TCTCCACTATATTGAAAGACTCCCAACATATTGGGCATTCCAGGCCTTCTTCCCTGCTCG 274

Qy         55 LysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySer 74
                    ||||||                          ||||||   |||||||||
Db        273 TGCACGAGGATGCTTCGTCATCGGAACAAGCAGCACTCGAAGAGGTGCAATTTGGGGGAC 214

Qy         75 SerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSerTyrLeuThrArg 94
              |||:::          :::|||   |||::::::|||   |||:::||||||
Db        213 TCCTCTTCCCCAGCCCGGATGCCCAAGCATTTGATGCAAAACCCCACATATCTGGAAGCT 154

Qy         95 IleGlnValThrLeuLeu 100
              |||      ||||||:::
Db        153 ATCTTCGAGACCCTGATT 136
```

```
RESULT 3
CF919839/c
LOCUS       CF919839                686 bp    mRNA    linear   EST 08-NOV-2003
DEFINITION  EST4139 Zea mays sperm cell cDNA library Zea mays cDNA clone
            Zmsp9923 5', mRNA sequence.
ACCESSION   CF919839
VERSION     CF919839.1  GI:38229526
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 686)
  AUTHORS   Engel,M.L., Chaboud,A., Dumas,C. and McCormick,S.
  TITLE     Sperm cells of Zea mays have a complex complement of mRNAs
  JOURNAL   Plant J. 34 (5), 697-707 (2003)
  MEDLINE   22674159
   PUBMED   12787250
COMMENT     Contact: McCormick, S.
            Plant Gene Expression Center
            USDA/ARS
            800 Buchanan St. Albany, CA 94710, USA
            Tel: 510 559 5906
            Fax: 510 559 5678
            Email: sheilamc@nature.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .686
                     /organism="Zea mays"
                     /mol_type="mRNA"
```

/cultivar="A188"
/db_xref="taxon:4577"
/clone="Zmsp9923"
/cell_type="sperm"
/clone_lib="Zea mays sperm cell cDNA library"
/note="Vector: Clontech Triplex2; RNA was isolated from fluorescent-activated cell sorted sperm cells. A cDNA library was constructed by oligo dT priming using the Clontech SMART cDNA synthesis kit. The cDNAs were directionally cloned."
ORIGIN

Alignment Scores:

Pred. No.:	6.62	Length:	686
Score:	86.50	Matches:	28
Percent Similarity:	46.51%	Conservative:	12
Best Local Similarity:	32.56%	Mismatches:	45
Query Match:	15.81%	Indels:	1
DB:	7	Gaps:	1

US-10-679-865A-2 (1-108) x CF919839 (1-686)

```
Qy         16 GluIleLeu---GluSerGluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGly 34
              |||:::|||      |||   |||                   |||   |||   :::
Db        458 GAAGTCCTAGGACGCAGTTCTTGCAAAGGGTGTGGCCGCACCATAGCACGTAAGGCACAT 399

Qy         35 SerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsn 54
              |||   :::|||   :::           |||   :::      ||||||      :::
Db        398 TCTCCACTATATTGAAAGACTCCCAACATATTGGGCATTCCAGGCCTTCTTCCCTGCTCG 339

Qy         55 LysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySer 74
                 ||||||                    ||||||   |||||||||
Db        338 TGCACGAGGATGCTTCGTCATCGGAACAAGCAGCACTCGAAGAGGTGCAATTTGGGGGAC 279

Qy         75 SerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSerTyrLeuThrArg 94
              |||:::            :::|||   |||::::::|||   |||:::||||||
Db        278 TCCTCTTCCCCAGCCCGGATGCCCAAGCATTTGATGCAAAACCCCACATATCTGGAAGCT 219

Qy         95 IleGlnValThrLeuLeu 100
              |||      ||||||:::
Db        218 ATCTTTAAGACCCTGATT 201
```

RESULT 4
CC332766/c
LOCUS CC332766 721 bp DNA linear GSS 16-MAY-2003
DEFINITION OGOAU48TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0341G23, genomic survey sequence.
ACCESSION CC332766
VERSION CC332766.1 GI:30801937
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 721)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .721
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMa0341G23"
/clone_lib="ZM_0.7_1.5_KB"
/note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"
ORIGIN

Alignment Scores:

Pred. No.:	7.11	Length:	721
Score:	86.50	Matches:	28
Percent Similarity:	46.51%	Conservative:	12
Best Local Similarity:	32.56%	Mismatches:	45
Query Match:	15.81%	Indels:	1
DB:	8	Gaps:	1

US-10-679-865A-2 (1-108) x CC332766 (1-721)

```
Qy         16 GluIleLeu---GluSerGluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGly 34
              |||:::|||      |||   |||                   |||   |||   :::
Db        624 GAAGTCCTAGGACGCAGTTCTTGCAAAGGGTGTGGCCGCACCATAGCACGTAAGGCACAT 565

Qy         35 SerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsn 54
              |||   :::|||   :::           |||   :::      ||||||      :::
Db        564 TCTCCACTATATTGAAAGACTCCCAACATATTGGGCATTCCAGGCCTTCTTCCCTGCTCG 505

Qy         55 LysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySer 74
                 ||||||                    ||||||   |||||||||
Db        504 TGCACGAGGATGCTTCGTCATCGGAACAAGCAGCACTCGAAGAGGTGCAATTTGGGGGAC 445

Qy         75 SerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSerTyrLeuThrArg 94
              |||:::            :::|||   |||::::::|||   |||:::||||||
Db        444 TCCTCTTCCCCAGCCCGGATGCCCAAGCATTTGATGCAAAACCCCACATATCTGGAAGCT 385

Qy         95 IleGlnValThrLeuLeu 100
              |||      ||||||:::
Db        384 ATCTTCAAGACCCTGATT 367
```

RESULT 5
CC149584
LOCUS CC149584 797 bp DNA linear GSS 24-JUN-2003
DEFINITION ZMMBBb0016F21.r ZMMBBb Zea mays genomic clone ZMMBBb0016F21 3', genomic survey sequence.
ACCESSION CC149584
VERSION CC149584.1 GI:30094777
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 797)
AUTHORS Yu,Y., Kim,H.R., Hatfield,J., Soderlund,C., Bharti,A.K., Messing,J. and Wing,R.
TITLE Sequencing of the maize genome
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: T7
BACKWARD: M13r
Plate: 0016 row: F column: 21
Seq primer: M13r
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .797
/organism="Zea mays"

```
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBBb0016F21"
                     /lab_host="DH10B"
                     /clone_lib="ZMMBBb"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Zea mays L. ssp. mays"
ORIGIN

Alignment Scores:
Pred. No.:              8.2           Length:        797
Score:                  86.50         Matches:       28
Percent Similarity:     46.51%        Conservative:  12
Best Local Similarity:  32.56%        Mismatches:    45
Query Match:            15.81%        Indels:        1
DB:                     8             Gaps:          1

US-10-679-865A-2 (1-108) x CC149584 (1-797)

Qy         16 GluIleLeu---GluSerGluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGly 34
              |||:::|||      |||   |||                 |||   |||   :::
Db        366 GAAGTCCTAGGACGCAGTTCTTGCAAAGGGTGTGGCCGCACCATAGCACGTAAGGCACAT 425

Qy         35 SerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsn 54
              |||   :::|||   :::             |||   :::      ||||||      :::
Db        426 TCTCCACTATATTGAAAGACTCCCAACATATTGGGCATTCCAGGCCTTCTTCCCTGCTCG 485

Qy         55 LysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySer 74
                    ||||||                          ||||||   |||||||||
Db        486 TGCACGAGGATGCTTCGTCATCGGAACAAGCAGCACTCGAAGAGGTGCAATTTGGGGGAC 545

Qy         75 SerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSerTyrLeuThrArg 94
              |||:::          :::|||   |||::::::|||   |||:::||||||
Db        546 TCCTCTTCCCCAGCCCGGATGCCCAAGCATTTGATGCAAAACCCCACATATCTGGAAGCT 605

Qy         95 IleGlnValThrLeuLeu 100
              |||      ||||||:::
Db        606 ATCTTCAAGACCCTGATT 623

RESULT 6
CC673519/c
LOCUS       CC673519                859 bp    DNA     linear   GSS 19-JUN-2003
DEFINITION  OGVEL95TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0536P22,
            genomic survey sequence.
ACCESSION   CC673519
VERSION     CC673519.1  GI:32077968
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 859)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OGVEL95TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .859
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0536P22"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

Alignment Scores:
Pred. No.:              9.13          Length:        859
Score:                  86.50         Matches:       28
Percent Similarity:     46.51%        Conservative:  12
Best Local Similarity:  32.56%        Mismatches:    45
Query Match:            15.81%        Indels:        1
DB:                     9             Gaps:          1

US-10-679-865A-2 (1-108) x CC673519 (1-859)

Qy         16 GluIleLeu---GluSerGluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGly 34
              |||:::|||      |||   |||                 |||   |||   :::
Db        365 GAAGTCCTAGGACGCAGTTCTTGCAAAGGGTGTGGCCGCACCATAGCACGTAAGGCACAT 306

Qy         35 SerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsn 54
              |||   :::|||   :::             |||   :::      ||||||      :::
Db        305 TCTCCACTATATTGAAAGACTCCCAACATATTGGGCATTCCAGGCCTTCTTCCCTGCTCG 246

Qy         55 LysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySer 74
                    ||||||                          ||||||   |||||||||
Db        245 TGCACGAGGATGCTTCGTCATCGGAACAAGCAGCACTCGAAGAGGTGCAATTTGGGGGAC 186

Qy         75 SerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSerTyrLeuThrArg 94
              |||:::          :::|||   |||::::::|||   |||:::||||||
Db        185 TCCTCTTCCCCAGCCCGGATGCCCAAGCATTTGATGCAAAACCCCACATATCTGGAAGCT 126

Qy         95 IleGlnValThrLeuLeu 100
              |||      ||||||:::
Db        125 ATCTTCAAGACCCTGATT 108

RESULT 7
BB757526/c
LOCUS       BB757526                530 bp    mRNA    linear   EST 17-OCT-2001
DEFINITION  BB757526 RIKEN full-length enriched, melanocyte Mus musculus cDNA
            clone G270085O21 3', mRNA sequence.
ACCESSION   BB757526
VERSION     BB757526.1  GI:16203746
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
            Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
            Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
```

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .530
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G270085021"
/cell_type="melanocyte"
/clone_lib="RIKEN full-length enriched, melanocyte"

ORIGIN

Alignment Scores:

Pred. No.:	5.26	Length:	530
Score:	86.00	Matches:	24
Percent Similarity:	49.33%	Conservative:	13
Best Local Similarity:	32.00%	Mismatches:	34
Query Match:	15.72%	Indels:	4
DB:	2	Gaps:	2

US-10-679-865A-2 (1-108) x BB757526 (1-530)

```
Qy          34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
               ||||||         |||||||||   |||       ||||||   |||:::
Db         464 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGCGGCCATAGCAGCCTGC 405

Qy          52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                  :::||||||   :::      ::::::::::::::
Db         404 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 345

Qy          71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                  :::|||:::   |||            |||:::|||   ||||||
Db         344 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 285

Qy          91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
               ||||||         |||||||||   |||:::|||:::
Db         284 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 240
```

RESULT 8
CR758635
LOCUS CR758635 313 bp mRNA linear EST 03-SEP-2004
DEFINITION CR758635 mdx substracted from control Mus musculus cDNA clone P0AB001ZE02, mRNA sequence.
ACCESSION CR758635
VERSION CR758635.1 GI:51879119
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 313)
AUTHORS Rouger,K., LeCunff,M., Steenman,M., Potier,M.C., Gibelin,N., Dechesne,C.A. and Leger,J.J.
TITLE Global/temporal gene expression in diaphragm and hindlimb muscles of dystrophin-deficient (mdx) mice
JOURNAL Am. J. Physiol. Cell Physiol. 283 (C), 773-784 (2002)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Genoscope sequence ID : P0AB001ZE02F1.

FEATURES Location/Qualifiers
source 1. .313
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="P0AB001ZE02"
/sex="male"
/tissue_type="muscle"
/clone_lib="mdx substracted from control"
/note="Vector: pCRII-TOPO; Pietu G., Cros N., Leger J.J., Dechesne C.A. Mus musculus substracted libraries, detail of cDNA collection: pool of 4 cDNA libraries: C57BL/10mdx diaphragm cDNA subtracted from C57BL/10ScSn diaphragm cDNA, C57BL/10ScSn diaphragm cDNA subtracted from C57BL/10mdx diaphragm cDNA, C57BL/10mdx hindlimb muscle cDNA substracted from C57BL/10ScSn hindlimb muscle cDNA, C57BL/10ScSn hindlimb muscle cDNA substracted from C57BL/10mdx hindlimb muscle cDNA."

ORIGIN

Alignment Scores:

Pred. No.:	3.28	Length:	313
Score:	85.00	Matches:	24
Percent Similarity:	49.33%	Conservative:	13
Best Local Similarity:	32.00%	Mismatches:	34
Query Match:	15.54%	Indels:	4
DB:	7	Gaps:	2

US-10-679-865A-2 (1-108) x CR758635 (1-313)

```
Qy          34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
               ||||||         |||||||||   |||       ||||||   |||:::
Db          73 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCCATAGCAGCCTGC 132

Qy          52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                  :::||||||   :::      ::::::::::::::
Db         133 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 192

Qy          71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                  :::|||:::   |||            |||:::|||   ||||||
Db         193 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 252

Qy          91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
               ||||||         |||||||||   |||:::|||:::
Db         253 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 297
```

RESULT 9
CR759545/c
LOCUS CR759545 313 bp mRNA linear EST 03-SEP-2004
DEFINITION CR759545 mdx substracted from control Mus musculus cDNA clone P0AB001ZE02, mRNA sequence.
ACCESSION CR759545
VERSION CR759545.1 GI:51880027
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 313)
AUTHORS Rouger,K., LeCunff,M., Steenman,M., Potier,M.C., Gibelin,N., Dechesne,C.A. and Leger,J.J.
TITLE Global/temporal gene expression in diaphragm and hindlimb muscles of dystrophin-deficient (mdx) mice
JOURNAL Am. J. Physiol. Cell Physiol. 283 (C), 773-784 (2002)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Genoscope sequence ID : P0AB001ZE02R1.

```
FEATURES             Location/Qualifiers
     source          1. .313
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="P0AB001ZE02"
                     /sex="male"
                     /tissue_type="muscle"
                     /clone_lib="mdx substracted from control"
                     /note="Vector: pCRII-TOPO; Pietu G., Cros N., Leger J.J.,
                     Dechesne C.A. Mus musculus substracted libraries, detail
                     of cDNA collection: pool of 4 cDNA libraries: C57BL/10mdx
                     diaphragm cDNA subtracted from C57BL/10ScSn diaphragm
                     cDNA, C57BL/10ScSn diaphragm cDNA subtracted from
                     C57BL/10mdx diaphragm cDNA, C57BL/10mdx hindlimb muscle
                     cDNA substracted from C57BL/10ScSn hindlimb muscle cDNA,
                     C57BL/10ScSn hindlimb muscle cDNA substracted from
                     C57BL/10mdx hindlimb muscle cDNA."
ORIGIN

Alignment Scores:
Pred. No.:                3.28          Length:          313
Score:                    85.00         Matches:         24
Percent Similarity:       49.33%        Conservative:    13
Best Local Similarity:    32.00%        Mismatches:      34
Query Match:              15.54%        Indels:          4
DB:                       7             Gaps:            2

US-10-679-865A-2 (1-108) x CR759545 (1-313)

Qy          34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
               ||||||          |||||||||   |||       ||||||   |||:::
Db         241 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCCATAGCAGCCTGC 182

Qy          52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                  :::||||||   :::      ::::::::::::::
Db         181 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 122

Qy          71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                  :::|||:::   |||          |||:::|||   ||||||
Db         121 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 62

Qy          91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
               ||||||          |||||||||   |||:::|||:::
Db          61 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 17

RESULT 10
BY585586/c
LOCUS       BY585586               382 bp    mRNA    linear   EST 15-DEC-2002
DEFINITION  BY585586 RIKEN full-length enriched, bone marrow mast cells Mus
            musculus cDNA clone K1C0024M05 3', mRNA sequence.
ACCESSION   BY585586
VERSION     BY585586.1  GI:26920768
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Tissues were provided by Hiromitsu Nakauchi ( Dept. of Immunology
            Institute of Basic Medical Sciences University of Tsukuba 1-1-1
            Tennodai,Tsukuba,Ibaraki 305-8578 ) whose assistance we gratefully
            acknowledge.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .382
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K1C0024M05"
                     /tissue_type="bone marrow"
                     /cell_type="mast cells"
                     /clone_lib="RIKEN full-length enriched, bone marrow mast
                     cells"
ORIGIN

Alignment Scores:
Pred. No.:                4.36          Length:          382
Score:                    85.00         Matches:         24
Percent Similarity:       49.33%        Conservative:    13
Best Local Similarity:    32.00%        Mismatches:      34
Query Match:              15.54%        Indels:          4
```

```
DB:                    6              Gaps:          2

US-10-679-865A-2 (1-108) x BY585586 (1-382)

Qy          34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
               ||||||         |||||||||   |||       ||||||   |||:::
Db         316 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCCATAGCAGCCTGC 257

Qy          52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                     :::||||||   :::      ::::::::::::::
Db         256 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 197

Qy          71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                  :::|||:::   |||            |||:::|||   ||||||
Db         196 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 137

Qy          91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
                  ||||||         |||||||||   |||:::|||:::
Db         136 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 92

RESULT 11
BF455298
LOCUS       BF455298                391 bp    mRNA    linear   EST 01-DEC-2000
DEFINITION  UI-M-BZ1-bkf-b-11-0-UI.s1 NIH_BMAP_MHI2_S1 Mus musculus cDNA clone
            UI-M-BZ1-bkf-b-11-0-UI 3', mRNA sequence.
ACCESSION   BF455298
VERSION     BF455298.1  GI:11521467
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 391)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            hippocampus tissue cDNA Library Preparation: M.B. Soares Lab Clone
            distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
            GENETICS. It should be noted that Bento Soares is generating a
            small number of additional specialized non-redundant arrays of BMAP
            cDNAs whose availability will be considered under appropriate and
            limited collaborative arrangements The following repetitive
            elements were found in this cDNA sequence: 1-21,
            >AT_rich#Low_complexity 336-403, >(GA)n#Simple_repeat 349-401,
            >(CAGA)n#Simple_repeat
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES            Location/Qualifiers
     source         1. .391
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="UI-M-BZ1-bkf-b-11-0-UI"
                    /dev_stage="27-32 days"
                    /lab_host="DH10B (Life Technologies)"
                    /clone_lib="NIH_BMAP_MHI2_S1"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; The
                    NIH_BMAP_MHI2_S1 library is a subtracted library derived
                    from  NIH_BMAP_MHI2. NIH_BMAP_MHI2 is a library derived
                    from mouse hippocampus tissue. For a detailed description
                    of the library from which this clone was  derived, please
                    visit our web site at brainest.eng.uiowa.edu.
                    TAG_TISSUE=hippocampus
                    TAG_LIB=NIH_BMAP_MHI2_S1
                    TAG_SEQ=TAGTC"
ORIGIN

Alignment Scores:
Pred. No.:              4.51          Length:          391
Score:                  85.00         Matches:         24
Percent Similarity:     49.33%        Conservative:    13
Best Local Similarity:  32.00%        Mismatches:      34
Query Match:            15.54%        Indels:          4
DB:                     2             Gaps:            2

US-10-679-865A-2 (1-108) x BF455298 (1-391)

Qy          34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
               ||||||         |||||||||   |||       ||||||   |||:::
Db          43 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCCATAGCAGCCTGC 102

Qy          52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                     :::||||||   :::      ::::::::::::::
Db         103 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 162

Qy          71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                  :::|||:::   |||            |||:::|||   ||||||
Db         163 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 222

Qy          91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
                  ||||||         |||||||||   |||:::|||:::
Db         223 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 267

RESULT 12
BY443486/c
LOCUS       BY443486                429 bp    mRNA    linear   EST 13-DEC-2002
DEFINITION  BY443486 RIKEN full-length enriched, 17 days embryo heart Mus
            musculus cDNA clone I920187E23 3', mRNA sequence.
ACCESSION   BY443486
VERSION     BY443486.1  GI:26730530
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
```

Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .429
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="I920187E23"
                     /tissue_type="heart"
                     /dev_stage="17 days embryo"
                     /clone_lib="RIKEN full-length enriched, 17 days embryo
                     heart"
ORIGIN

Alignment Scores:
Pred. No.:              5.15          Length:        429
Score:                  85.00         Matches:       24
Percent Similarity:     49.33%        Conservative:  13
Best Local Similarity:  32.00%        Mismatches:    34
Query Match:            15.54%        Indels:        4
DB:                     5             Gaps:          2

US-10-679-865A-2 (1-108) x BY443486 (1-429)

Qy          34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
               ||||||         |||||||||   |||      ||||||   |||:::
Db         349 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCCATAGCAGCCTGC 290

Qy          52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                      :::||||||   :::      :::::::::::::::
Db         289 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 230

Qy          71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                          :::|||:::   |||            |||:::|||   ||||||
Db         229 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 170

Qy          91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
                  ||||||           |||||||||   |||:::|||:::
Db         169 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 125

RESULT 13
AA000493
LOCUS       AA000493                445 bp    mRNA    linear   EST 18-JUL-1996
DEFINITION  mg23b04.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:424591 5', mRNA sequence.
ACCESSION   AA000493
VERSION     AA000493.1  GI:1436357
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:259143
            Seq primer: ETPrimer
            High quality sequence stop: 330.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:424591"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:              5.43          Length:        445
Score:                  85.00         Matches:       24
Percent Similarity:     49.33%        Conservative:  13
Best Local Similarity:  32.00%        Mismatches:    34
Query Match:            15.54%        Indels:        4
DB:                     1             Gaps:          2
```

```
US-10-679-865A-2 (1-108) x AA000493 (1-445)

Qy          34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
               ||||||         |||||||||   |||       ||||||   |||:::
Db          75 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCCATAGCAGCCTGC 134

Qy          52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                  :::||||||   :::      ::::::::::::::
Db         135 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 194

Qy          71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                  :::|||:::   |||             |||:::|||   ||||||
Db         195 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 254

Qy          91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
               ||||||         |||||||||   |||:::|||:::
Db         255 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 299

RESULT 14
BM939154/c
LOCUS       BM939154                453 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  UI-M-BZ1-bkf-b-11-0-UI.r1 NIH_BMAP_MHI2_S1 Mus musculus cDNA clone
            UI-M-BZ1-bkf-b-11-0-UI 5', mRNA sequence.
ACCESSION   BM939154
VERSION     BM939154.1  GI:19398306
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 41-108, >(GA)n#Simple_repeat (matched compliment) 43-95,
            >(CAGA)n#Simple_repeat (matched compliment) 426-451,
            >POLY_A#Simple_repeat
            Seq primer: M13 REVERSE.
FEATURES             Location/Qualifiers
     source          1. .453
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BZ1-bkf-b-11-0-UI"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NIH_BMAP_MHI2_S1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_MHI2_S1 library is a subtracted library derived
                     from  NIH_BMAP_MHI2. NIH_BMAP_MHI2 is a library derived
                     from mouse hippocampus tissue. For a detailed description
                     of the library from which this clone was  derived, please
                     visit our web site at brainest.eng.uiowa.edu."
ORIGIN

Alignment Scores:
Pred. No.:              5.57          Length:          453
Score:                  85.00         Matches:         24
Percent Similarity:     49.33%        Conservative:    13
Best Local Similarity:  32.00%        Mismatches:      34
Query Match:            15.54%        Indels:          4
DB:                     5             Gaps:            2

US-10-679-865A-2 (1-108) x BM939154 (1-453)

Qy          34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
               ||||||         |||||||||   |||       ||||||   |||:::
Db         352 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCCATAGCAGCCTGC 293

Qy          52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                  :::||||||   :::      ::::::::::::::
Db         292 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 233

Qy          71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                  :::|||:::   |||             |||:::|||   ||||||
Db         232 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 173

Qy          91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
               ||||||         |||||||||   |||:::|||:::
Db         172 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 128

RESULT 15
BE449045
LOCUS       BE449045                454 bp    mRNA    linear   EST 25-JUL-2000
DEFINITION  ut85g04.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:3369270 5', mRNA sequence.
ACCESSION   BE449045
VERSION     BE449045.1  GI:9448622
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 454)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1078874
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .454
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3369270"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:              5.59          Length:          454
Score:                  85.00         Matches:         24
Percent Similarity:     49.33%        Conservative:    13
```

```
Best Local Similarity:  32.00%          Mismatches:    34
Query Match:            15.54%          Indels:        4
DB:                     2               Gaps:          2

US-10-679-865A-2 (1-108) x BE449045 (1-454)

Qy         34 GlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu------ 51
              ||||||         |||||||||   |||      ||||||   |||:::
Db         81 GGCTCCCCACAGGCACCAGAGGCTACTGCGGCCAGGCCCTTGGTGGCCATAGCAGCCTGC 140

Qy         52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
                     :::||||||   :::      ::::::::::::::
Db        141 TCTTCTGAGCAAAGGGAACCCACTATGGAGACCCAAAGTATGATGGCATTTTCAGGGCAG 200

Qy         71 AsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSer 90
                    :::|||:::   |||           |||:::|||   ||||||
Db        201 GAGGCAAAAGAGGCCGCCCTCCCCAGGAGAAAGGAAAGACACTTGTTGGTTAGGCCTGAA 260

Qy         91 TyrLeuThrArg---IleGlnValThrLeuLeuCysAsnSerSer 104
              ||||||         |||||||||   |||:::|||:::
Db        261 GTCTTAACAATGGGAACATAAGTTACGCTGTTCTGTTCTAGCACT 305


Search completed: September  1, 2005, 00:09:01
Job time : 2665 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 31, 2005, 11:39:04 ; Search time 446 Seconds
(without alignments)
1433.480 Million cell updates/sec

Title: US-10-679-865A-2
Perfect score: 547
Sequence: 1 MVTLLPIEGQEVHFFEILES.....PSYLTRIQVTLLCNSSAEAL 108

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US10679865/runat_30082005_150800_15905/app_query.fasta_1.263
-DB=N_Geneseq -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10679865_@CGN_1_1_470_@runat_30082005_150800_15905 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	547	100.0	862	8	ABS56399	Abs56399 Human BRC
c	2	78.5	14.4	26042	8	AAD56102	Aad56102 Mouse Ccn
c	3	78.5	14.4	26042	9	ADA02464	Ada02464 Mouse Ccn
c	4	78.5	14.4	26042	10	ADB72203	Adb72203 Mouse Ccn
	5	77.5	14.2	777	8	ACA38357	Aca38357 Prokaryot
	6	77.5	14.2	69350	12	ADM98959	Adm98959 Diterpene
	7	77.5	14.2	110000	4	AAI99682_37	Continuation (38 o
	8	77.5	14.2	110000	4	AAI99682_38	Continuation (39 o
	9	77.5	14.2	110000	4	AAI99683_37	Continuation (38 o
c	10	77	14.1	45190	6	ABX09144	Abx09144 Mycobacte
c	11	77	14.1	110000	4	AAI99682_11	Continuation (12 o
c	12	77	14.1	110000	4	AAI99683_11	Continuation (12 o
	13	75.5	13.8	1094	5	AAS69164	Aas69164 DNA encod
	14	74.5	13.6	466	9	ACH45637	Ach45637 Human foe
c	15	74	13.5	1340	3	AAZ90452	Aaz90452 Wnt-4AF a
c	16	74	13.5	1682	11	ADM03332	Adm03332 Human cDN
c	17	74	13.5	2057	2	AAX19936	Aax19936 Human Wnt
c	18	74	13.5	2184	13	ACN39347	Acn39347 Tumour-as
c	19	74	13.5	2190	2	AAX19935	Aax19935 Human Wnt
c	20	74	13.5	2251	12	ADO22229	Ado22229 Human WNT
c	21	74	13.5	2251	13	ACN39348	Acn39348 Tumour-as
c	22	74	13.5	2252	6	ABK92158	Abk92158 Prostate
c	23	74	13.5	2252	11	ADN39265	Adn39265 Cancer/an
	24	74	13.5	12494	10	ADE53916	Ade53916 Human pro
	25	74	13.5	12632	8	ACA03928	Aca03928 cDNA down
c	26	74	13.5	76201	12	ADJ62814	Adj62814 Human cDN
	27	72.5	13.3	1134	5	AAS67967	Aas67967 DNA encod
	28	72.5	13.3	4437	12	ADJ75924	Adj75924 Marker ge
c	29	72.5	13.3	10872	10	ABT42282	Abt42282 Toxicity
c	30	72.5	13.3	11272	2	AAX89842	Aax89842 Rat cubil
c	31	72.5	13.3	11272	12	ADP71882	Adp71882 Renal tox
	32	72.5	13.3	12747	10	ADF81724	Adf81724 Leukaemia
	33	72.5	13.3	12747	10	ADF81723	Adf81723 Leukaemia
c	34	72	13.2	441	4	AAS05974	Aas05974 Human bet
c	35	72	13.2	525	10	ACF16983	Acf16983 Human NOV
	36	72	13.2	705	8	ACA40818	Aca40818 Prokaryot
	37	72	13.2	1053	12	ADN13936	Adn13936 Human pro
	38	72	13.2	3960	13	ADR06784	Adr06784 Full leng
	39	72	13.2	6144	13	ACN42337	Acn42337 Human dia
	40	72	13.2	6487	12	ADL16583	Adl16583 Human 282
	41	72	13.2	6487	12	ADL16581	Adl16581 Human 282
	42	72	13.2	6487	12	ADL16430	Adl16430 Human 282
	43	72	13.2	7329	12	ADL16587	Adl16587 Human 282
	44	72	13.2	7329	12	ADL16432	Adl16432 Human 282
	45	72	13.2	7329	12	ADL16589	Adl16589 Human 282

ALIGNMENTS

RESULT 1
ABS56399
ID ABS56399 standard; cDNA; 862 BP.
XX
AC ABS56399;
XX
DT 24-JAN-2003 (first entry)
XX
DE Human BRCC2 cDNA.
XX
KW Human; gene; ss; BRCC2; apoptosis; cancer; proliferation; apoptosis;
KW metastasis; inhibitor; ribozyme; antisense; gene therapy; radiotherapy;
KW chemotherapy; anticancer agent; hormone; cell cycle dependent kinase;
KW antisense raf; tumour; therapeutic; Raf-1; chromosome 11q; lung cancer;
KW breast cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 294. .620
FT /*tag= a
FT /product= "BRCC2"
XX
PN WO200281639-A2.
XX
PD 17-OCT-2002.
XX
PF 08-APR-2002; 2002WO-US010848.
XX

PR 06-APR-2001; 2001US-0281796P.
XX
PA (GEOU) UNIV GEORGETOWN.
XX
PI Kasid U, Gokhale P, Kumar D, Broustas C, Ahmad I, Dritschilo A;
PI Rahman A;
XX
DR WPI; 2003-058521/05.
DR P-PSDB; ABG72006.
XX
PT Novel isolated BRCC2 polypeptide which modulates apoptosis and nucleic
PT acid encoding the polypeptide, useful for treating conditions
PT characterized by BRCC2 underexpression, and as diagnostic target for
PT detecting cancer.
XX
PS Claim 2; Fig 1; 48pp; English.
XX
CC The invention discloses an isolated polypeptide, and polynucleotide
CC encoding it, of the human BRCC2 (not defined) protein. The BRCC2
CC polypeptide modulates apoptosis. The polynucleotide, polypeptide and
CC antibodies raised against the polypeptide, can be used for treating or
CC inhibiting cancer cell proliferation, apoptosis and/or metastasis in a
CC cancer, of preferably the breast or lung, patient by administering a
CC small molecule inhibitor of BRCC2, an antibody that specifically binds
CC BRCC2, a ribozyme or antisense oligonucleotide (gene therapy) that
CC modulates BRCC2 expression which can be used in combination with
CC radiotherapy, chemotherapy, biological anticancer agent, hormones or
CC inhibitors of cell cycle dependent kinases. They are also useful for
CC detecting, treating and for evaluating the prognosis of conditions,
CC particularly cancer, characterised by BRCC2 over or underexpression, for
CC identifying compounds that inhibit cancer progression by modulating
CC apoptosis of cancer and as markers for normal physiological function
CC related to cell growth or survival. BRCC2s advantage is that decreasing
CC the amount of BRCC2 induces many of the in vivo effects of antisense raf
CC oligonucleotide, such as tumour growth arrest, tumour regression, tumour
CC cell death and/or potentiate radiation/drug-induced cytotoxicity. In
CC addition, being a therapeutic target potentially downstream of Raf-1,
CC greater specificity is obtained by targeting the action of BRCC2 as
CC compared to Raf-1. The sequence presented is the human BRCC2 cDNA and its
CC gene is located on chromosome 11q
XX
SQ Sequence 862 BP; 272 A; 163 C; 185 G; 242 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.56e-58	Length:	862
Score:	547.00	Matches:	108
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-10-679-865A-2 (1-108) x ABS56399 (1-862)

```
Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCT 353

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        354 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 413

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        414 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 473

Qy         61 LysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        474 AAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAAT 533

Qy         81 ValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnValThrLeuLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        534 GTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTA 593

Qy        101 CysAsnSerSerAlaGluAlaLeu 108
              |||||||||||||||||||||||||
Db        594 TGCAATTCCTCTGCTGAGGCCCTG 617
```

RESULT 2
AAD56102/c
ID AAD56102 standard; DNA; 26042 BP.
XX
AC AAD56102;
XX
DT 07-AUG-2003 (first entry)
XX
DE Mouse Ccnd3 carcinoma associated (CA) gene.
XX
KW Carcinoma; gene therapy; carcinoma associated gene; CA; Ccnd3; mouse; ds.
XX
OS Mus sp.
XX
PN WO2003035837-A2.
XX
PD 01-MAY-2003.
XX
PF 22-OCT-2002; 2002WO-US033835.
XX
PR 23-OCT-2001; 2001US-00004113.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Engelhard EK, Morris DW;
XX
DR WPI; 2003-421412/39.
XX
PT New recombinant nucleic acid and its encoded protein, useful for
PT preparing a composition for diagnosing or treating carcinomas.
XX
PS Claim 1; Page 99-102; 173pp; English.
XX
CC The invention relates to novel sequences which are useful for preparing a
CC composition for diagnosing or treating carcinomas. These sequence are
CC also useful in gene therapy. The present sequence is mouse Ccnd3
CC carcinoma associated (CA) gene. This sequence is used in the invention
XX
SQ Sequence 26042 BP; 6194 A; 6390 C; 6396 G; 6669 T; 0 U; 393 Other;

Alignment Scores:

Pred. No.:	823	Length:	26042
Score:	78.50	Matches:	25
Percent Similarity:	54.17%	Conservative:	14
Best Local Similarity:	34.72%	Mismatches:	26
Query Match:	14.35%	Indels:	7
DB:	8	Gaps:	2

US-10-679-865A-2 (1-108) x AAD56102 (1-26042)

```
Qy         30 GluArgAlaSerGlySerSerIleTyrProGluAla--------------LysAlaArg 44
              :::|||   ||||||||||:::   ||||||   |||                  |||
Db      22203 CAAAGGCTCTCGGGCAGCACCTCATACCCAGTGGCTCTTGATCTGACAAACTTAGCAGAA 22144

Qy         45 LeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuProLysGluThrVal 64
                 |||:::   |||   ||||||:::         |||:::   :::   ||||||
Db      22143 TGTTTTCTGCAGTTGCTATACGGAAGTGACAAC------ATGGTGGAAAGGGGGACCGTG 22090

Qy         65 LeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsnValProGlyHis 84
              :::|||::::::::::   |||   ||||||      :::      |||   |||:::
Db      22089 GGCACTCTTGAGCAGTGGCTGCTGGCATCTTCAGAAACCAGGCTAAGGGTCTTTGGGAAC 22030

Qy         85 ValLeuGlnArgProSerTyrLeuThrArgIleGln 96
                    ||||||   :::|||       |||:::
Db      22029 ACAGCAATCAGGCCTTGTCACTTAAGGCACATTGAA 21994
```

RESULT 3
ADA02464/c
ID ADA02464 standard; DNA; 26042 BP.

```
XX
AC   ADA02464;
XX
DT   06-NOV-2003  (first entry)
XX
DE   Mouse Ccnd3 carcinoma associated gene, SEQ ID NO:983.
XX
KW   Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW   prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW   gene; ds.
XX
OS   Mus sp.
XX
PN   WO2003057146-A2.
XX
PD   17-JUL-2003.
XX
PF   26-DEC-2002; 2002WO-US041414.
XX
PR   26-DEC-2001; 2001US-00035832.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-587068/55.
XX
PT   New recombinant nucleic acid encoding carcinoma associated protein,
PT   useful for preparing compositions for treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 983; 245pp; English.
XX
CC   The invention relates to recombinant carcinoma associated (CA) nucleic
CC   acid sequences from mouse and human (ADA01482-ADA03094), and to
CC   recombinant carcinoma associated proteins (CAP) encoded by them. The
CC   invention also encompasses expression vectors and host cells comprising a
CC   CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC   binds to the protein, and a biochip comprising CA nucleic acid or
CC   fragments thereof. The sequences of the invention were identified using
CC   oncogenic retroviruses, which insert into the genome of the host organism
CC   at random. Many of these do not carry transduced host oncogenes or
CC   pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC   direct consequence of the effects of proviral integration into host
CC   protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC   carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC   leukaemia) or a propensity to carcinoma by determination of the sequence
CC   of a CA gene, or by determination of CA gene expression in particular
CC   tissues. CA nucleic acids, proteins and antibodies are also useful as
CC   therapeutic agents and in screening and evaluating drug candidates. The
CC   present sequence represents a specifically claimed murine CA nucleic acid
CC   sequence of the invention. Note: The complete sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 26042 BP; 6194 A; 6390 C; 6396 G; 6669 T; 0 U; 393 Other;

Alignment Scores:
Pred. No.:              823           Length:        26042
Score:                  78.50         Matches:       25
Percent Similarity:     54.17%        Conservative:  14
Best Local Similarity:  34.72%        Mismatches:    26
Query Match:            14.35%        Indels:        7
DB:                     9             Gaps:          2

US-10-679-865A-2 (1-108) x ADA02464 (1-26042)

Qy         30 GluArgAlaSerGlySerSerIleTyrProGluAla--------------LysAlaArg 44
              :::|||   |||||||||:::   ||||||   |||                  |||
Db      22203 CAAAGGCTCTCGGGCAGCACCTCATACCCAGTGGCTCTTGATCTGACAAACTTAGCAGAA 22144

Qy         45 LeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuProLysGluThrVal 64
                    |||:::   |||   ||||||:::         |||:::   :::   ||||||
```

```
Db      22143 TGTTTTCTGCAGTTGCTATACGGAAGTGACAAC------ATGGTGGAAAGGGGGACCGTG 22090

Qy         65 LeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsnValProGlyHis 84
              :::|||::::::::::   |||   ||||||    :::      |||   |||:::
Db      22089 GGCACTCTTGAGCAGTGGCTGCTGGCATCTTCAGAAACCAGGCTAAGGGTCTTTGGGAAC 22030

Qy         85 ValLeuGlnArgProSerTyrLeuThrArgIleGln 96
                    ||||||   :::|||      |||:::
Db      22029 ACAGCAATCAGGCCTTGTCACTTAAGGCACATTGAA 21994

RESULT 4
ADB72203/c
ID   ADB72203 standard; DNA; 26042 BP.
XX
AC   ADB72203;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Mouse Ccnd3 gene.
XX
KW   mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW   cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS   Mus sp.
XX
PN   WO2003008583-A2.
XX
PD   30-JAN-2003.
XX
PF   26-DEC-2001; 2001WO-US051291.
XX
PR   02-MAR-2001; 2001US-00798586.
PR   23-OCT-2001; 2001US-00004113.
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-00997722.
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-239337/23.
XX
PT   New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT   cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS   Claim 1; SEQ ID NO 31; 2304pp; English.
XX
CC   The invention relates to a novel recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the 660 sequences fully defined
CC   in the specification. A polynucleotide of the invention has cytostatic
CC   activity, and may have a use in gene therapy, or in a vaccine. The
CC   recombinant nucleic acids and polypeptides are useful for treating
CC   carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC   sarcomas. The present sequence represents a mouse gene of the invention.
XX
SQ   Sequence 26042 BP; 6194 A; 6390 C; 6396 G; 6669 T; 0 U; 393 Other;

Alignment Scores:
Pred. No.:              823           Length:        26042
Score:                  78.50         Matches:       25
Percent Similarity:     54.17%        Conservative:  14
Best Local Similarity:  34.72%        Mismatches:    26
Query Match:            14.35%        Indels:        7
DB:                     10            Gaps:          2

US-10-679-865A-2 (1-108) x ADB72203 (1-26042)

Qy         30 GluArgAlaSerGlySerSerIleTyrProGluAla--------------LysAlaArg 44
              :::|||   |||||||||:::   ||||||   |||                  |||
Db      22203 CAAAGGCTCTCGGGCAGCACCTCATACCCAGTGGCTCTTGATCTGACAAACTTAGCAGAA 22144

Qy         45 LeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuProLysGluThrVal 64
```

```
                    |||:::   |||   ||||||:::        |||:::   :::   ||||||
Db      22143 TGTTTTCTGCAGTTGCTATACGGAAGTGACAAC------ATGGTGGAAAGGGGGACCGTG 22090

Qy         65 LeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsnValProGlyHis 84
                 :::|||:::::::::   |||   ||||||      :::      |||   |||:::
Db      22089 GGCACTCTTGAGCAGTGGCTGCTGGCATCTTCAGAAACCAGGCTAAGGGTCTTTGGGAAC 22030

Qy         85 ValLeuGlnArgProSerTyrLeuThrArgIleGln 96
                      ||||||   :::|||     |||:::
Db      22029 ACAGCAATCAGGCCTTGTCACTTAAGGCACATTGAA 21994

RESULT 5
ACA38357
ID   ACA38357 standard; DNA; 777 BP.
XX
AC   ACA38357;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Prokaryotic essential gene #20014.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Mycobacterium bovis.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU34487.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 26227; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC   prokaryotic essential genes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 777 BP; 113 A; 223 C; 286 G; 155 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                8.53          Length:          777
Score:                    77.50         Matches:         34
Percent Similarity:       35.04%        Conservative:    14
Best Local Similarity:    24.82%        Mismatches:      46
Query Match:              14.17%        Indels:          43
DB:                       8             Gaps:            5

US-10-679-865A-2 (1-108) x ACA38357 (1-777)

Qy          5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
              |||      ||||||   |||       |||      :::      :::      |||
Db         49 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTGTAGGCGATCAGGCATCGTTG 108

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
              :::                              |||:::   ||||||          |||
Db        109 TGG-------------------------GAATCCGTGCTGCCCGAGGAGTTGCGGCGG 141

Qy         45 LeuPro---------------LeuGluAlaLeuLeuGlySerAsnLys------------ 55
              ||||||               ::::::|||||||||      :::
Db        142 CTGCCCGAAGAGCTGGCCCGGGTGGATGCGCTGCTCGATGATTCGGCGTTCTTCTGCCCG 201

Qy         56 ----------------------------GluProMetLeuProLysGluThrValLeu 65
                                          |||   :::|||   ||||||   |||
Db        202 TTTGTGCCGTTCTTCGACCCGCGGATGGGTCGGCCGTCCATACCGATGGAGACCTATTTG 261

Qy         66 SerLeuLys------------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
              |||               ||||||   ||||||      ::::::   |||   |||
Db        262 CGGTTGATGTTCTTGAAGTTCCGTTACCGGTTGGGCTATGAGTCGCTGTGTCGGGAGGTC 321

Qy         82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                     :::      |||                                   |||
Db        322 ACCGATTCGATCACCTGGCGGCGGTTCTGCCGTATTCCGTTGGAGGGATCGGTGCCGCAC 381

Qy         91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAlaGluAla 107
                 |||   :::::::::|||      |||      |||   |||
Db        382 CCAACCACGTTGATGAAGCTGACCACGCGCTGCGGTGAGGATGCGGTGGCC 432

RESULT 6
ADM98959
ID   ADM98959 standard; DNA; 69350 BP.
XX
AC   ADM98959;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Diterpene synthase DNA #19.
XX
KW   Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;
KW   diterpene precursor; diterpene synthase; defence toxin;
KW   volatile defensive signal; pollinator attractant; photoprotectant.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US2004072323-A1.
XX
PD   15-APR-2004.
XX
PF   07-JAN-2002; 2002US-00041018.
XX
```

```
PR   05-JAN-2001; 2001US-0259880P.
XX
PA   (MATS/) MATSUDA S P T.
PA   (HART/) HART E A.
XX
PI   Matsuda SPT,  Hart EA;
XX
DR   WPI; 2004-373921/35.
XX
PT   New unicellular organisms comprising exogenous nucleic acids encoding a
PT   geranylgeranyl pyrophosphate and a diterpene synthase, useful for
PT   producing diterpenes and diterpene precursors.
XX
PS   Claim 3; SEQ ID NO 379; 38pp; English.
XX
CC   The invention relates to a unicellular organism for producing a diterpene
CC   or diterpene precursor comprising an exogenous nucleic acid sequence
CC   encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC   promoter operable in the organism, and an exogenous nucleic acid sequence
CC   encoding a diterpene synthase under the control of a promoter operable in
CC   the organism. The invention also relates to methods of producing a
CC   diterpene or diterpene precursor and a method of isolating a diterpene
CC   synthase comprising growing several cells in the presence of a
CC   polyaromatic resin to make a cell/resin mixture, where at least one of
CC   the cells further comprises at least one isolated and purified nucleic
CC   acid sequence of a yeast expression library, and the expression of the
CC   nucleic acid sequence is regulated by an inducible promoter under
CC   conditions where the expression is induced, filtering the cell/resin
CC   mixture, extracting the cell/resin mixture with alcohol to produce an
CC   organic eluent and analysing the organic eluent by a screening method
CC   including chromatography and/or spectroscopy, to identify the nucleic
CC   acid sequence encoding the diterpene synthase. The unicellular
CC   microorganism is useful as a diterpene or diterpene precursor producing
CC   system. Diterpenes, in plants, serve as defence toxins, volatile
CC   defensive signals, pollinator attractants and photoprotectants. This
CC   sequence represents DNA encoding a diterpene synthase polypeptide of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the printed specification but was obtained in electronic format from
CC   USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 69350 BP; 12311 A; 22891 C; 22715 G; 11433 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.25e+03        Length:         69350
Score:                  77.50           Matches:        34
Percent Similarity:     35.04%          Conservative:   14
Best Local Similarity:  24.82%          Mismatches:     46
Query Match:            14.17%          Indels:         43
DB:                     12              Gaps:           5

US-10-679-865A-2 (1-108) x ADM98959 (1-69350)

Qy          5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
              |||      ||||||   |||         |||      :::      :::      |||
Db      57011 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTGTAGGCGATCAGGCATCGTTG 57070

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
              :::                        |||:::   ||||||      |||
Db      57071 TGG-------------------------GAATCCGTGCTGCCCGAGGAGTTGCGGCGG 57103

Qy         45 LeuPro--------------LeuGluAlaLeuLeuGlySerAsnLys------------ 55
              ||||||              ::::::|||||||||      :::
Db      57104 CTGCCCGAAGAGCTGGCCCGGGTGGATGCGCTGCTCGATGATTCGGCGTTCTTCTGCCCG 57163

Qy         56 ------------------------------GluProMetLeuProLysGluThrValLeu 65
                                           |||   :::|||   ||||||   |||
Db      57164 TTTGTGCCGTTCTTCGACCCGCGGATGGGTCGGCCGTCCATACCGATGGAGACCTATTTG 57223

Qy         66 SerLeuLys------------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
                 |||              ||||||   ||||||      ::::::   |||   |||
Db      57224 CGGTTGATGTTCTTGAAGTTCCGTTACCGGTTGGGCTATGAGTCGCTGTGTCGGGAGGTC 57283

Qy         82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                              :::      |||                              |||
Db      57284 ACCGATTCGATCACCTGGCGGCGGTTCTGCCGTATTCCGTTGGAGGGATCGGTGCCGCAC 57343

Qy         91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAlaGluAla 107
                    |||   :::::::::|||      |||         |||   |||
Db      57344 CCAACCACGTTGATGAAGCTGACCACGCGCTGCGGTGAGGATGCGGTGGCC 57394

RESULT 7
AAI99682_37
Continuation (38 of 45) of AAI99682 from base 3700001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments  LOCUS AAI99682 Accession Aai99682
WP      Fragment Name       Begin      End
WP      AAI99682_00               1    110000
WP      AAI99682_01          100001    210000
WP      AAI99682_02          200001    310000
WP      AAI99682_03          300001    410000
WP      AAI99682_04          400001    510000
WP      AAI99682_05          500001    610000
WP      AAI99682_06          600001    710000
WP      AAI99682_07          700001    810000
WP      AAI99682_08          800001    910000
WP      AAI99682_09          900001   1010000
WP      AAI99682_10         1000001   1110000
WP      AAI99682_11         1100001   1210000
WP      AAI99682_12         1200001   1310000
WP      AAI99682_13         1300001   1410000
WP      AAI99682_14         1400001   1510000
WP      AAI99682_15         1500001   1610000
WP      AAI99682_16         1600001   1710000
WP      AAI99682_17         1700001   1810000
WP      AAI99682_18         1800001   1910000
WP      AAI99682_19         1900001   2010000
WP      AAI99682_20         2000001   2110000
WP      AAI99682_21         2100001   2210000
WP      AAI99682_22         2200001   2310000
WP      AAI99682_23         2300001   2410000
WP      AAI99682_24         2400001   2510000
WP      AAI99682_25         2500001   2610000
WP      AAI99682_26         2600001   2710000
WP      AAI99682_27         2700001   2810000
WP      AAI99682_28         2800001   2910000
WP      AAI99682_29         2900001   3010000
WP      AAI99682_30         3000001   3110000
WP      AAI99682_31         3100001   3210000
WP      AAI99682_32         3200001   3310000
WP      AAI99682_33         3300001   3410000
WP      AAI99682_34         3400001   3510000
WP      AAI99682_35         3500001   3610000
WP      AAI99682_36         3600001   3710000
WP      AAI99682_37         3700001   3810000
WP      AAI99682_38         3800001   3910000
WP      AAI99682_39         3900001   4010000
WP      AAI99682_40         4000001   4110000
WP      AAI99682_41         4100001   4210000
WP      AAI99682_42         4200001   4310000
WP      AAI99682_43         4300001   4410000
WP      AAI99682_44         4400001   4411529

Alignment Scores:
Pred. No.:              8.05e+03        Length:         110000
Score:                  77.50           Matches:        34
Percent Similarity:     35.04%          Conservative:   14
Best Local Similarity:  24.82%          Mismatches:     46
Query Match:            14.17%          Indels:         43
DB:                     4               Gaps:           5

US-10-679-865A-2 (1-108) x AAI99682_37 (1-110000)

Qy          5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
              |||      ||||||   |||         |||      :::      :::      |||
Db     100061 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTGTAGGCGATCAGGCATCGTTG 100120

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
```

```
              :::                                 |||:::   ||||||         |||
Db     100121 TGG--------------------------GAATCCGTGCTGCCCGAGGAGTTGCGGCGG 100153

Qy         45 LeuPro--------------LeuGluAlaLeuLeuGlySerAsnLys------------ 55
              ||||||               ::::::|||||||||      :::
Db     100154 CTGCCCGAAGAGCTGGCCCGGGTGGATGCGCTGCTCGATGATTCGGCGTTCTTCTGCCCG 100213

Qy         56 ----------------------------GluProMetLeuProLysGluThrValLeu 65
                                           |||   :::|||   ||||||   |||
Db     100214 TTTGTGCCGTTCTTCGACCCGCGGATGGGTCGGCCGTCCATACCGATGGAGACCTATTTG 100273

Qy         66 SerLeuLys------------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
              |||               ||||||  ||||||      ::::::   |||   |||
Db     100274 CGGTTGATGTTCTTGAAGTTCCGTTACCGGTTGGGCTATGAGTCGCTGTGTCGGGAGGTC 100333

Qy         82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                        :::      |||                             |||
Db     100334 ACCGATTCGATCACCTGGCGGCGGTTCTGCCGTATTCCGTTGGAGGGATCGGTGCCGCAC 100393

Qy         91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAlaGluAla 107
                    |||   :::::::::|||      |||          |||   |||
Db     100394 CCAACCACGTTGATGAAGCTGACCACGCGCTGCGGTGAGGATGCGGTGGCC 100444

RESULT 8
AAI99682_38
Continuation (39 of 45) of AAI99682 from base 3800001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments  LOCUS AAI99682 Accession Aai99682
WP     Fragment Name      Begin     End
WP     AAI99682_00            1    110000
WP     AAI99682_01       100001    210000
WP     AAI99682_02       200001    310000
WP     AAI99682_03       300001    410000
WP     AAI99682_04       400001    510000
WP     AAI99682_05       500001    610000
WP     AAI99682_06       600001    710000
WP     AAI99682_07       700001    810000
WP     AAI99682_08       800001    910000
WP     AAI99682_09       900001   1010000
WP     AAI99682_10      1000001   1110000
WP     AAI99682_11      1100001   1210000
WP     AAI99682_12      1200001   1310000
WP     AAI99682_13      1300001   1410000
WP     AAI99682_14      1400001   1510000
WP     AAI99682_15      1500001   1610000
WP     AAI99682_16      1600001   1710000
WP     AAI99682_17      1700001   1810000
WP     AAI99682_18      1800001   1910000
WP     AAI99682_19      1900001   2010000
WP     AAI99682_20      2000001   2110000
WP     AAI99682_21      2100001   2210000
WP     AAI99682_22      2200001   2310000
WP     AAI99682_23      2300001   2410000
WP     AAI99682_24      2400001   2510000
WP     AAI99682_25      2500001   2610000
WP     AAI99682_26      2600001   2710000
WP     AAI99682_27      2700001   2810000
WP     AAI99682_28      2800001   2910000
WP     AAI99682_29      2900001   3010000
WP     AAI99682_30      3000001   3110000
WP     AAI99682_31      3100001   3210000
WP     AAI99682_32      3200001   3310000
WP     AAI99682_33      3300001   3410000
WP     AAI99682_34      3400001   3510000
WP     AAI99682_35      3500001   3610000
WP     AAI99682_36      3600001   3710000
WP     AAI99682_37      3700001   3810000
WP     AAI99682_38      3800001   3910000
WP     AAI99682_39      3900001   4010000
WP     AAI99682_40      4000001   4110000
WP     AAI99682_41      4100001   4210000
WP     AAI99682_42      4200001   4310000
WP     AAI99682_43      4300001   4410000
WP     AAI99682_44      4400001   4411529
```

```
Alignment Scores:
Pred. No.:                8.05e+03        Length:          110000
Score:                    77.50           Matches:         34
Percent Similarity:       35.04%          Conservative:    14
Best Local Similarity:    24.82%          Mismatches:      46
Query Match:              14.17%          Indels:          43
DB:                       4               Gaps:            5

US-10-679-865A-2 (1-108) x AAI99682_38 (1-110000)

Qy          5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
              |||      ||||||   |||       |||      :::      :::      |||
Db         61 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTGTAGGCGATCAGGCATCGTTG 120

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
              :::                              |||:::   ||||||         |||
Db        121 TGG--------------------------GAATCCGTGCTGCCCGAGGAGTTGCGGCGG 153

Qy         45 LeuPro--------------LeuGluAlaLeuLeuGlySerAsnLys------------ 55
              ||||||               ::::::|||||||||      :::
Db        154 CTGCCCGAAGAGCTGGCCCGGGTGGATGCGCTGCTCGATGATTCGGCGTTCTTCTGCCCG 213

Qy         56 ----------------------------GluProMetLeuProLysGluThrValLeu 65
                                           |||   :::|||   ||||||   |||
Db        214 TTTGTGCCGTTCTTCGACCCGCGGATGGGTCGGCCGTCCATACCGATGGAGACCTATTTG 273

Qy         66 SerLeuLys------------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
              |||               ||||||  ||||||      ::::::   |||   |||
Db        274 CGGTTGATGTTCTTGAAGTTCCGTTACCGGTTGGGCTATGAGTCGCTGTGTCGGGAGGTC 333

Qy         82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                        :::      |||                             |||
Db        334 ACCGATTCGATCACCTGGCGGCGGTTCTGCCGTATTCCGTTGGAGGGATCGGTGCCGCAC 393

Qy         91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAlaGluAla 107
                    |||   :::::::::|||      |||          |||   |||
Db        394 CCAACCACGTTGATGAAGCTGACCACGCGCTGCGGTGAGGATGCGGTGGCC 444

RESULT 9
AAI99683_37
Continuation (38 of 44) of AAI99683 from base 3700001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments  LOCUS AAI99683 Accession Aai99683
WP     Fragment Name      Begin     End
WP     AAI99683_00            1    110000
WP     AAI99683_01       100001    210000
WP     AAI99683_02       200001    310000
WP     AAI99683_03       300001    410000
WP     AAI99683_04       400001    510000
WP     AAI99683_05       500001    610000
WP     AAI99683_06       600001    710000
WP     AAI99683_07       700001    810000
WP     AAI99683_08       800001    910000
WP     AAI99683_09       900001   1010000
WP     AAI99683_10      1000001   1110000
WP     AAI99683_11      1100001   1210000
WP     AAI99683_12      1200001   1310000
WP     AAI99683_13      1300001   1410000
WP     AAI99683_14      1400001   1510000
WP     AAI99683_15      1500001   1610000
WP     AAI99683_16      1600001   1710000
WP     AAI99683_17      1700001   1810000
WP     AAI99683_18      1800001   1910000
WP     AAI99683_19      1900001   2010000
WP     AAI99683_20      2000001   2110000
WP     AAI99683_21      2100001   2210000
WP     AAI99683_22      2200001   2310000
WP     AAI99683_23      2300001   2410000
WP     AAI99683_24      2400001   2510000
WP     AAI99683_25      2500001   2610000
WP     AAI99683_26      2600001   2710000
WP     AAI99683_27      2700001   2810000
WP     AAI99683_28      2800001   2910000
```

```
WP      AAI99683_29       2900001   3010000
WP      AAI99683_30       3000001   3110000
WP      AAI99683_31       3100001   3210000
WP      AAI99683_32       3200001   3310000
WP      AAI99683_33       3300001   3410000
WP      AAI99683_34       3400001   3510000
WP      AAI99683_35       3500001   3610000
WP      AAI99683_36       3600001   3710000
WP      AAI99683_37       3700001   3810000
WP      AAI99683_38       3800001   3910000
WP      AAI99683_39       3900001   4010000
WP      AAI99683_40       4000001   4110000
WP      AAI99683_41       4100001   4210000
WP      AAI99683_42       4200001   4310000
WP      AAI99683_43       4300001   4403765

Alignment Scores:
Pred. No.:                8.05e+03        Length:         110000
Score:                    77.50           Matches:        34
Percent Similarity:       35.04%          Conservative:   14
Best Local Similarity:    24.82%          Mismatches:     46
Query Match:              14.17%          Indels:         43
DB:                       4               Gaps:           5

US-10-679-865A-2 (1-108) x AAI99683_37 (1-110000)

Qy          5 LeuProIleGluGlyGlnGluValHisPheGluIleLeuGluSerGluCysValLeu 24
              |||      ||||||   |||         |||      :::     :::       |||
Db      90883 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTGTAGGCGATCAGGCATCGTTG 90942

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
              :::                              |||:::   ||||||         |||
Db      90943 TGG-------------------------GAATCCGTGCTGCCCGAGGAGTTGCGGCGG 90975

Qy         45 LeuPro--------------LeuGluAlaLeuLeuGlySerAsnLys----------- 55
              ||||||              ::::::|||||||||     :::
Db      90976 CTGCCCGAAGAGCTGGCCCGGGTGGATGCGCTGCTCGATGATTCGGCGTTCTTCTGCCCG 91035

Qy         56 ----------------------------GluProMetLeuProLysGluThrValLeu 65
                                          |||   :::|||   ||||||   |||
Db      91036 TTTGTGCCGTTCTTCGACCCGCGGATGGGTCGGCCGTCCATACCGATGGAGACCTATTTG 91095

Qy         66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
              |||               ||||||   ||||||      ::::::   |||   |||
Db      91096 CGGTTGATGTTCTTGAAGTTCCGTTACCGGTTGGGCTATGAGTCGCTGTGTCGGGAGGTC 91155

Qy         82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                    :::      |||                                    |||
Db      91156 ACCGATTCGATCACCTGGCGGCGGTTCTGCCGTATTCCGTTGGAGGGATCGGTGCCGCAC 91215

Qy         91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAlaGluAla 107
                 |||   :::::::::|||       |||         |||   |||
Db      91216 CCAACCACGTTGATGAAGCTGACCACGCGCTGCGGTGAGGATGCGGTGGCC 91266

RESULT 10
ABX09144/c
ID   ABX09144 standard; DNA; 45190 BP.
XX
AC   ABX09144;
XX
DT   08-APR-2003  (first entry)
XX
DE   Mycobacterium tuberculosis H37Rv BAC clone BAC-Rv267.
XX
KW   Mycobacterioses; survival; virulence; protective antigen; vaccine;
KW   mycobacterial disease; tuberculosis; leprosy; ds; cosmid.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO200274903-A2.
XX
PD   26-SEP-2002.
XX
PF   22-FEB-2002; 2002WO-IB001973.
XX
PR   22-FEB-2001; 2001US-0270123P.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Cole S;
XX
DR   WPI; 2002-759885/82.
XX
PT   Identifying and selecting genes for survival or virulence of mycobacteria
PT   by a comparative genomic analysis of the sequences of Mycobacterium
PT   tuberculosis and M. leprae.
XX
PS   Disclosure; Fig 8; 874pp; English.
XX
CC   This invention relates to a novel method for identifying essential genes
CC   for survival or virulence of mycobacteria species. The method comprises
CC   aligning the genomic sequence of a first mycobacterium species on a
CC   genomic sequence of a second mycobacterium species and selecting a
CC   polynucleotide sequence that is highly conserved in both genomes with no
CC   counterparts in other bacterial genomic sequences and that corresponds to
CC   an essential gene for the survival or virulence of mycobacterium species.
CC   The method of the invention is useful for detecting M. tuberculosis or M.
CC   leprae infection. The method reduces the number of potential new targets
CC   and protective antigens for new drugs and vaccine compositions to treat
CC   and prevent mycobacterial diseases, particularly tuberculosis and
CC   leprosy. The present sequence represents a Mycobacterial cosmid DNA
CC   sequence used in the method of the invention
XX
SQ   Sequence 45190 BP; 7490 A; 14522 C; 15204 G; 7974 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                2.72e+03        Length:         45190
Score:                    77.00           Matches:        37
Percent Similarity:       34.97%          Conservative:   13
Best Local Similarity:    25.87%          Mismatches:     45
Query Match:              14.08%          Indels:         48
DB:                       6               Gaps:           6

US-10-679-865A-2 (1-108) x ABX09144 (1-45190)

Qy          5 LeuProIleGluGlyGlnGluValHisPheGluIleLeuGluSerGluCysValLeu 24
              |||      ||||||   |||         |||      |||     :::|||   |||
Db      35763 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTTTAGGCGATCAGTGCTGGTTG 35704

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
              :::                              |||:::   |||:::         |||
Db      35703 TGG-------------------------GAGTCGCTGCTGCCCGACGAGGTGCGCCGA 35671

Qy         45 LeuPro--------------LeuGluAlaLeuLeuGlySer------------------ 53
              ||||||              ::::::|||||||||
Db      35670 CTGCCCGAGGAACTGGCCCGGGTGGACGCATTGTTGGACGATCCGGCGTTCTTCGCCCCG 35611

Qy         54 ------------------------AsnLysGluProMetLeuProLysGluThrValLeu 65
                                       |||      |||   |||      |||
Db      35610 TTCGTGCCGTTCTTCGACCCGCGCAGGGGCCGGCCGTCGACGCCGATGGAGGTCTATCTG 35551

Qy         66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
                 |||               ||||||   ||||||      ::::::   |||   |||
Db      35550 CAGTTGATGTTTGTGAAGTTCCGCTACCGGCTGGGCTATGAGTCGCTGTGCCGGGAGGTG 35491

Qy         82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                    :::      |||                                    |||
Db      35490 GCTGATTCGATCACCTGACGGCGGTTTTGCCGCATTGCGCTGGACGGGTCGGTGCCGCAT 35431

Qy       . .91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAla--------------- 105
                 |||   :::::::::|||       |||   |||:::|||
Db      35430 CCGACCACATTGATGAAGCTCACCACGCGTTGCGGATCGGCGGCGATCGACGGGCTCAAC 35371

Qy        106 GluAlaLeu 108
                |||||||||
Db      35370 GAGGCATTG 35362
```

```
RESULT 11
AAI99682_11/c
Continuation (12 of 45) of AAI99682 from base 1100001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments  LOCUS AAI99682 Accession Aai99682
WP      Fragment Name       Begin      End
WP      AAI99682_00             1    110000
WP      AAI99682_01        100001    210000
WP      AAI99682_02        200001    310000
WP      AAI99682_03        300001    410000
WP      AAI99682_04        400001    510000
WP      AAI99682_05        500001    610000
WP      AAI99682_06        600001    710000
WP      AAI99682_07        700001    810000
WP      AAI99682_08        800001    910000
WP      AAI99682_09        900001   1010000
WP      AAI99682_10       1000001   1110000
WP      AAI99682_11       1100001   1210000
WP      AAI99682_12       1200001   1310000
WP      AAI99682_13       1300001   1410000
WP      AAI99682_14       1400001   1510000
WP      AAI99682_15       1500001   1610000
WP      AAI99682_16       1600001   1710000
WP      AAI99682_17       1700001   1810000
WP      AAI99682_18       1800001   1910000
WP      AAI99682_19       1900001   2010000
WP      AAI99682_20       2000001   2110000
WP      AAI99682_21       2100001   2210000
WP      AAI99682_22       2200001   2310000
WP      AAI99682_23       2300001   2410000
WP      AAI99682_24       2400001   2510000
WP      AAI99682_25       2500001   2610000
WP      AAI99682_26       2600001   2710000
WP      AAI99682_27       2700001   2810000
WP      AAI99682_28       2800001   2910000
WP      AAI99682_29       2900001   3010000
WP      AAI99682_30       3000001   3110000
WP      AAI99682_31       3100001   3210000
WP      AAI99682_32       3200001   3310000
WP      AAI99682_33       3300001   3410000
WP      AAI99682_34       3400001   3510000
WP      AAI99682_35       3500001   3610000
WP      AAI99682_36       3600001   3710000
WP      AAI99682_37       3700001   3810000
WP      AAI99682_38       3800001   3910000
WP      AAI99682_39       3900001   4010000
WP      AAI99682_40       4000001   4110000
WP      AAI99682_41       4100001   4210000
WP      AAI99682_42       4200001   4310000
WP      AAI99682_43       4300001   4410000
WP      AAI99682_44       4400001   4411529

Alignment Scores:
Pred. No.:              9.3e+03        Length:         110000
Score:                  77.00          Matches:        37
Percent Similarity:     34.97%         Conservative:   13
Best Local Similarity:  25.87%         Mismatches:     45
Query Match:            14.08%         Indels:         48
DB:                     4              Gaps:           6

US-10-679-865A-2 (1-108) x AAI99682_11 (1-110000)

Qy          5 LeuProIleGluGlyGlnGluValHisPheGluIleLeuGluSerGluCysValLeu 24
              |||      ||||||   |||       |||      |||      :::|||   |||
Db      60383 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTTTAGGCGATCAGTGCTGGTTG 60324

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
              :::                          |||:::   |||:::         |||
Db      60323 TGG-------------------------GAGTCGCTGCTGCCCGACGAGGTGCGCCGA 60291

Qy         45 LeuPro--------------LeuGluAlaLeuLeuGlySer----------------- 53
              ||||||              ::::::|||||||||
Db      60290 CTGCCCGAGGAACTGGCCCGGGTGGACGCATTGTTGGACGATCCGGCGTTCTTCGCCCCG 60231

Qy         54 ----------------------AsnLysGluProMetLeuProLysGluThrValLeu 65
                                     |||      |||   |||      |||
Db      60230 TTCGTGCCGTTCTTCGACCCGCGCAGGGGCCGGCCGTCGACGCCGATGGAGGTCTATCTG 60171

Qy         66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
                 |||              ||||||   |||||      ::::::   |||   |||
Db      60170 CAGTTGATGTTTGTGAAGTTCCGCTACCGGCTGGGCTATGAGTCGCTGTGCCGGGAGGTG 60111

Qy         82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                      :::      |||                                  |||
Db      60110 GCTGATTCGATCACCTGACGGCGGTTTTGCCGCATTGCGCTGGACGGGTCGGTGCCGCAT 60051

Qy         91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAla--------------- 105
                    |||   :::::::::|||      |||   |||:::|||
Db      60050 CCGACCACATTGATGAAGCTCACCACGCGTTGCGGATCGGCGGCGATCGACGGGCTCAAC 59991

Qy        106 GluAlaLeu 108
                 |||||||||
Db      59990 GAGGCATTG 59982

RESULT 12
AAI99683_11/c
Continuation (12 of 44) of AAI99683 from base 1100001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments  LOCUS AAI99683 Accession Aai99683
WP      Fragment Name       Begin      End
WP      AAI99683_00             1    110000
WP      AAI99683_01        100001    210000
WP      AAI99683_02        200001    310000
WP      AAI99683_03        300001    410000
WP      AAI99683_04        400001    510000
WP      AAI99683_05        500001    610000
WP      AAI99683_06        600001    710000
WP      AAI99683_07        700001    810000
WP      AAI99683_08        800001    910000
WP      AAI99683_09        900001   1010000
WP      AAI99683_10       1000001   1110000
WP      AAI99683_11       1100001   1210000
WP      AAI99683_12       1200001   1310000
WP      AAI99683_13       1300001   1410000
WP      AAI99683_14       1400001   1510000
WP      AAI99683_15       1500001   1610000
WP      AAI99683_16       1600001   1710000
WP      AAI99683_17       1700001   1810000
WP      AAI99683_18       1800001   1910000
WP      AAI99683_19       1900001   2010000
WP      AAI99683_20       2000001   2110000
WP      AAI99683_21       2100001   2210000
WP      AAI99683_22       2200001   2310000
WP      AAI99683_23       2300001   2410000
WP      AAI99683_24       2400001   2510000
WP      AAI99683_25       2500001   2610000
WP      AAI99683_26       2600001   2710000
WP      AAI99683_27       2700001   2810000
WP      AAI99683_28       2800001   2910000
WP      AAI99683_29       2900001   3010000
WP      AAI99683_30       3000001   3110000
WP      AAI99683_31       3100001   3210000
WP      AAI99683_32       3200001   3310000
WP      AAI99683_33       3300001   3410000
WP      AAI99683_34       3400001   3510000
WP      AAI99683_35       3500001   3610000
WP      AAI99683_36       3600001   3710000
WP      AAI99683_37       3700001   3810000
WP      AAI99683_38       3800001   3910000
WP      AAI99683_39       3900001   4010000
WP      AAI99683_40       4000001   4110000
WP      AAI99683_41       4100001   4210000
WP      AAI99683_42       4200001   4310000
WP      AAI99683_43       4300001   4403765

Alignment Scores:
Pred. No.:              9.3e+03        Length:         110000
```

```
Score:                  77.00          Matches:       37
Percent Similarity:     34.97%         Conservative:  13
Best Local Similarity:  25.87%         Mismatches:    45
Query Match:            14.08%         Indels:        48
DB:                     4              Gaps:          6

US-10-679-865A-2 (1-108) x AAI99683_11 (1-110000)

Qy          5 LeuProIleGluGlyGlnGluValHisPheGluIleLeuGluSerGluCysValLeu 24
              |||      ||||||   |||        |||      |||     :::|||   |||
Db      60408 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTTTAGGCGATCAGTGCTGGTTG 60349

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
              :::                              |||:::   |||:::         |||
Db      60348 TGG-------------------------GAGTCGCTGCTGCCCGACGAGGTGCGCCGA 60316

Qy         45 LeuPro--------------LeuGluAlaLeuLeuGlySer----------------- 53
              ||||||              ::::::||||||||
Db      60315 CTGCCCGAGGAACTGGCCCGGGTGGACGCATTGTTGGACGATCCGGCGTTCTTCGCCCCG 60256

Qy         54 ----------------------AsnLysGluProMetLeuProLysGluThrValLeu 65
                                      |||      |||  |||      |||
Db      60255 TTCGTGCCGTTCTTCGACCCGCGCAGGGGCCGGCCGTCGACGCCGATGGAGGTCTATCTG 60196

Qy         66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
              |||           ||||||  ||||||     ::::::  |||  |||
Db      60195 CAGTTGATGTTTGTGAAGTTCCGCTACCGGCTGGGCTATGAGTCGCTGTGCCGGGAGGTG 60136

Qy         82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                 :::     |||                                     |||
Db      60135 GCTGATTCGATCACCTGACGGCGGTTTTGCCGCATTGCGCTGGACGGGTCGGTGCCGCAT 60076

Qy         91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAla-------------- 105
              |||  :::::::::|||     |||  |||:::|||
Db      60075 CCGACCACATTGATGAAGCTCACCACGCGTTGCGGATCGGCGGCGATCGACGGGCTCAAC 60016

Qy        106 GluAlaLeu 108
              |||||||||
Db      60015 GAGGCATTG 60007

RESULT 13
AAS69164
ID   AAS69164 standard; cDNA; 1094 BP.
XX
AC   AAS69164;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #4968.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG04977.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 4968; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1094 BP; 391 A; 268 C; 277 G; 158 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              24.4           Length:        1094
Score:                  75.50          Matches:       22
Percent Similarity:     46.15%         Conservative:  14
Best Local Similarity:  28.21%         Mismatches:    27
Query Match:            13.80%         Indels:        15
DB:                     5              Gaps:          3

US-10-679-865A-2 (1-108) x AAS69164 (1-1094)

Qy         31 ArgAlaSerGlySerSerIleTyrProGlu---AlaLysAlaArgLeuProLeuGluAla 49
              ||||||   |||:::      :::   :::   ||||||   :::|||      |||
Db        220 AGGGCAAAAGGCACGAAAGCATTTAAGCAACCTGCAAAGCGGAAGCTGGACGCAGAAGAT 279

Qy         50 LeuLeuGlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArg 69
              ::::::||||||   ::::::|||      |||||||||             |||:::
Db        280 GTAATTGGCAGCAGGAGACAGCCAAGAGCACCTAAGGAAAAGGCCCAACCCCTGGAA--- 336

Qy         70 TyrAsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgPro 89
                 :::|||   |||                  |||||||||      :::
Db        337 ---GATCTGGCCAGCTTCCAAGAGCTCTCTCAAACACCAGGCCACACTGAGGAA------ 387

Qy         90 SerTyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAlaGluAla 107
                                           |||   |||   :::|||::::::
Db        388 ----------------------------CTGGCAAATGGTGCTGCTGATAGC 411

RESULT 14
ACH45637
ID   ACH45637 standard; cDNA; 466 BP.
XX
AC   ACH45637;
XX
DT   13-OCT-2003  (first entry)
XX
DE   Human foetal brain cDNA #6362.
XX
KW   Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW   genome mapping; biodiversity; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   US2003073623-A1.
XX
PD   17-APR-2003.
```

```
XX
PF   30-JUL-2001; 2001US-00918995.
XX
PR   30-JUL-2001; 2001US-00918995.
XX
PA   (DRMA/) DRMANAC R T.
PA   (LABA/) LABAT I.
PA   (STAC/) STACHE-CRAIN B.
PA   (DICK/) DICKSON M C.
PA   (JONE/) JONES L W.
XX
PI   Drmanac RT,  Labat I,  Stache-Crain B,  Dickson MC,  Jones LW;
XX
DR   WPI; 2003-615964/58.
XX
PT   New polynucleotide sequences obtained from various cDNA libraries, useful
PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 32849; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversities, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping, in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present sequence
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ   Sequence 466 BP; 120 A; 155 C; 108 G; 82 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:              9.99          Length:        466
Score:                  74.50         Matches:       28
Percent Similarity:     47.37%        Conservative:  17
Best Local Similarity:  29.47%        Mismatches:    38
Query Match:            13.62%        Indels:        12
DB:                     9             Gaps:          4

US-10-679-865A-2 (1-108) x ACH45637 (1-466)

Qy         18 LeuGluSerGluCysValLeuTyrThrGlyTrpIleGluArgAlaSer-GlySerSerIl 37
              |||   ||||||   :::   :::   ||||||      |||   |||    :::|||
Db         12 CTGGCTTCTGAGGCCATTTCCTTCCGCGGCTGG------CGCCACAGCCACCACCAGCAC 65

Qy         37 eTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySer---------As 54
                                   |||       |||       ||||||
Db         66 GTCCTCCACGGTCACCGTCACGGCCCCGGCCCCCGCCGCCACAGGAAGCCCAGTGAAAAA 125

Qy         54 nLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySe 74
              :::   |||:::||||||||||||||||      ::::::::::|||           ||
Db        126 GCAGAGGCCGCTTTTACCGAAGGAGACTGCCCCGGCCGTGCAGCGGGTCGTGTGGAACTC 185

Qy         74 rSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSerTyrLeuThrAr 94
              ||||   :::::::::               :::   |||   ||||||      |||
Db        186 ATCAACTGTCCAGCAGAAG---------GAGATCACACAGAGCCCATCCACGTCCACC-- 234

Qy         94 gIleGlnValThrLeuLeuCysAsnSerSerAlaGluAlaLeu 108
                 :::||||||:::   ::::::   :::      |||
Db        235 -------ATCACCCTGGTGACCAGCACACAGTCATCGCCCCTG 270

RESULT 15

AAZ90452/c
ID   AAZ90452 standard; DNA; 1340 BP.
XX
AC   AAZ90452;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Wnt-4AF and Wnt-5c homolog polypeptide #2 encoding DNA.
XX
KW   Wnt; Wnt-4AF; Wnt-5C; frizzled gene; membrane protein; diagnostic;
KW   therapeutic; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63. .1142
FT                   /*tag=  a
XX
PN   WO200012117-A1.
XX
PD   09-MAR-2000.
XX
PF   20-AUG-1999;   99WO-US019046.
XX
PR   31-AUG-1998;   98US-0098440P.
PR   31-AUG-1998;   98US-0098453P.
PR   30-OCT-1998;   98US-0106462P.
PR   09-DEC-1998;   98US-0111588P.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Edmonds BT,  Su W;
XX
DR   WPI; 2000-256491/22.
DR   P-PSDB; AAY57271.
XX
PT   New gene encoding the Wnt-4AF and Wnt-5C homolog polypeptide involved in
PT   developmental control during embryonic development is useful to diagnose
PT   and treat related disease.
XX
PS   Claim 3; Page 73; 41pp; English.
XX
CC   The invention provides isolated nucleic acid sequences (AAZ90451-454)
CC   encoding Wnt-4AF and Wnt-5C homolog polypeptides (AAY57270-273). The
CC   polypeptides can be expressed by standard recombinant methodology. They
CC   upregulate gene expression by binding to Frizzled class of membrane
CC   proteins The Wnt antibodies can be used in diagnostics and therapeutics,
CC   while chimeric and transgenic animals can provide models of disease for
CC   testing the effectiveness of therapeutic or diagnostic agents
XX
SQ   Sequence 1340 BP; 283 A; 380 C; 430 G; 247 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              49.7          Length:        1340
Score:                  74.00         Matches:       31
Percent Similarity:     41.51%        Conservative:  13
Best Local Similarity:  29.25%        Mismatches:    38
Query Match:            13.53%        Indels:        24
DB:                     3             Gaps:          5

US-10-679-865A-2 (1-108) x AAZ90452 (1-1340)

Qy         12 ValHisPhePheGluIleLeuGlu----------------------------SerGlu 21
              |||||||||      :::   :::                            |||
Db       1111 GTGCACTTCTTCTACACCTGACGAAGCAGCACCAGTGGAACTTGCAGTGGCAGCGCTCCACC 1052

Qy         22 CysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAla 41
              |||   |||                 |||                 |||      |||
Db       1051 TGCACGCTCTTGAACTGGTTGTAGCCACGCCCGCAGCACATGAGCTCACAGCCATCC--- 995

Qy         42 LysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuProLys 61
                    :::|||   |||   ||||||         :::   |||   :::|||:::
Db        994 ---------ATGCCCTCCGAGGTCTTGTTG------CAGAGGCGGCCCTGCGTGCCCAGG 950
```

```
Qy          62 GluThrValLeuSerLeuLysArgTyr------AsnLeuGlySerSerAlaMetLysArg 79
               |||   |||||||||||:::|||           |||||||||:::     :::
Db         949 GAGCCCGTGCTCTCGTTGCGCAGGCAGTAGTCGGGGCTGGGGTCCACATAGACCAGGTCC 890

Qy          80 AsnValProGlyHisValLeuGln-----------------ArgProSerTyrLeuThr 93
               :::      |||   |||   :::                 ||||||      :::|||
Db         889 TCCGGGGTGGGCTGGGTGAAGCGGCTGTTGACCAGCTCCAGCCGGCCCTTGCGGGTGACG 830

Qy          94 ArgIleGlnValThrLeu 99
               |||:::         |||
Db         829 CGCATGGCGGCCGCGCTG 812
```

Search completed: August 31, 2005, 23:07:10
Job time : 556 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 31, 2005, 14:04:42 ; Search time 540 Seconds
(without alignments)
1308.618 Million cell updates/sec

Title: US-10-679-865A-2
Perfect score: 547
Sequence: 1 MVTLLPIEGQEVHFFEILES.....PSYLTRIQVTLLCNSSAEAL 108

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 7331713 seqs, 3271544945 residues

Total number of hits satisfying chosen parameters: 14663426

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US10679865/runat_30082005_150803_15929/app_query.fasta_1.263
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10679865_@CGN_1_1_480_@runat_30082005_150803_15929
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	547	100.0	862	19	US-10-679-865A-1	Sequence 1, Appli
2	80	14.6	1353	19	US-10-437-963-43632	Sequence 43632, A
c 3	78.5	14.4	26042	16	US-10-004-113-31	Sequence 31, Appl
4	77.5	14.2	777	17	US-10-282-122A-26227	Sequence 26227, A
5	77.5	14.2	69350	18	US-10-041-018-379	Sequence 379, App
c 6	77	14.1	45191	15	US-10-080-170-649	Sequence 649, App
c 7	77	14.1	45191	19	US-10-080-170-649	Sequence 649, App
c 8	77	14.1	45191	19	US-10-468-356-649	Sequence 649, App
9	76	13.9	1056	19	US-10-767-701-10785	Sequence 10785, A
10	74.5	13.6	466	10	US-09-918-995-32849	Sequence 32849, A
c 11	74.5	13.6	52128	20	US-10-719-993-7051	Sequence 7051, Ap
c 12	74	13.5	361	18	US-10-424-599-28834	Sequence 28834, A
c 13	74	13.5	1682	17	US-10-108-260A-2017	Sequence 2017, Ap
c 14	74	13.5	1926	13	US-10-027-632-98121	Sequence 98121, A
c 15	74	13.5	1926	13	US-10-027-632-98122	Sequence 98122, A
c 16	74	13.5	1926	13	US-10-027-632-98123	Sequence 98123, A
c 17	74	13.5	1926	13	US-10-027-632-100365	Sequence 100365,
c 18	74	13.5	1926	13	US-10-027-632-100366	Sequence 100366,
c 19	74	13.5	1926	17	US-10-027-632-98121	Sequence 98121, A
c 20	74	13.5	1926	17	US-10-027-632-98122	Sequence 98122, A
c 21	74	13.5	1926	17	US-10-027-632-98123	Sequence 98123, A
c 22	74	13.5	1926	17	US-10-027-632-100365	Sequence 100365,
c 23	74	13.5	1926	17	US-10-027-632-100366	Sequence 100366,
c 24	74	13.5	2251	16	US-10-285-976-16	Sequence 16, Appl
c 25	74	13.5	2252	17	US-10-295-027-583	Sequence 583, App
c 26	74	13.5	2252	21	US-10-847-972-15	Sequence 15, Appl
c 27	74	13.5	2826	22	US-10-852-707-2	Sequence 2, Appli
28	74	13.5	12494	16	US-10-252-157-263	Sequence 263, App
29	74	13.5	12632	13	US-10-071-766-49	Sequence 49, Appl
c 30	74	13.5	76201	17	US-10-439-703-84	Sequence 84, Appl
c 31	73.5	13.4	54648	17	US-10-085-117-289	Sequence 289, App
32	72.5	13.3	786	20	US-10-425-115-23230	Sequence 23230, A
33	72.5	13.3	826	20	US-10-425-115-113770	Sequence 113770,
c 34	72.5	13.3	10872	18	US-10-152-319A-1984	Sequence 1984, Ap
35	72	13.2	705	17	US-10-282-122A-28688	Sequence 28688, A
36	72	13.2	6487	18	US-10-435-751-6	Sequence 6, Appli
37	72	13.2	6487	18	US-10-435-751-157	Sequence 157, App
38	72	13.2	6487	18	US-10-435-751-159	Sequence 159, App
39	72	13.2	7329	18	US-10-435-751-8	Sequence 8, Appli
40	72	13.2	7329	18	US-10-435-751-163	Sequence 163, App
41	72	13.2	7329	18	US-10-435-751-165	Sequence 165, App
42	72	13.2	7491	18	US-10-435-751-4	Sequence 4, Appli
43	72	13.2	7491	18	US-10-435-751-151	Sequence 151, App
44	72	13.2	7491	18	US-10-435-751-153	Sequence 153, App
45	72	13.2	7527	18	US-10-435-751-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-10-679-865A-1
; Sequence 1, Application US/10679865A
; Publication No. US20040115714A1
; GENERAL INFORMATION:
; APPLICANT: NeoPharm, Inc.
; APPLICANT: Kasid, Usha
; APPLICANT: Gokhale, Prafulla
; APPLICANT: Kumar, Deepak
; APPLICANT: Broustas, Constantinos
; APPLICANT: Ahmad, Imran
; APPLICANT: Dritschilo, Anatoly
; APPLICANT: Rahman, Aquilur
; TITLE OF INVENTION: GENE BRCC-2 AND DIAGNOSITIC AND THERAPEUTIC USES THEREOF
; FILE REFERENCE: 224382
; CURRENT APPLICATION NUMBER: US/10/679,865A

```
;  CURRENT FILING DATE:  2003-10-03
;  PRIOR APPLICATION NUMBER: PCT/US02/10848
;  PRIOR FILING DATE: 2002-04-08
;  PRIOR APPLICATION NUMBER: 60/281,796
;  PRIOR FILING DATE: 2001-04-06
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 862
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (294)..(620)
US-10-679-865A-1

Alignment Scores:
Pred. No.:              5.07e-70      Length:        862
Score:                  547.00        Matches:       108
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     19            Gaps:          0

US-10-679-865A-2 (1-108) x US-10-679-865A-1 (1-862)

Qy          1 MetValThrLeuLeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 ATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCT 353

Qy         21 GluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        354 GAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAG 413

Qy         41 AlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        414 GCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCT 473

Qy         61 LysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        474 AAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAAT 533

Qy         81 ValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnValThrLeuLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        534 GTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTA 593

Qy        101 CysAsnSerSerAlaGluAlaLeu 108
              ||||||||||||||||||||||||
Db        594 TGCAATTCCTCTGCTGAGGCCCTG 617

RESULT 2
US-10-437-963-43632
; Sequence 43632, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 43632
;   LENGTH: 1353
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_46770C.1
US-10-437-963-43632

Alignment Scores:
Pred. No.:              0.727         Length:        1353
Score:                  80.00         Matches:       30
Percent Similarity:     45.74%        Conservative:  13
Best Local Similarity:  31.91%        Mismatches:    35
Query Match:            14.63%        Indels:        16
DB:                     19            Gaps:          4

US-10-679-865A-2 (1-108) x US-10-437-963-43632 (1-1353)

Qy         30 GluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAla 49
              ||||||||||||   :::   :::     :::|||   :::     |||   |||
Db        484 GAGCGCGCCTCCGTGGCGGGCGTG---GACGACGCGACGTCGGCCAAGCCGGAAGAGACG 540

Qy         50 LeuLeuGlySerAsnLysGluPro---------MetLeuProLysGluThrValLeuSer 66
                  |||          :::|||               |||:::|||      ||||||
Db        541 ACGTGCGGCGGCGTGCATGATCCGGCGAAGGCGAGCTCGCCGGAGGAGGCGGCCCTGAGC 600

Qy         67 ---LeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsnValProGlyHis--- 84
                 |||           |||         :::   ||||||   :::   |||
Db        601 ATCCTGCATTCGCTCAAGCTCTCCGAGGAGAGCTTCAAGAGGGTCTTGGAGGGCAGCGGC 660

Qy         85 ----------------------------ValLeuGlnArgProSerTyrLeuThrArg 94
                                          ||||||:::|||||||||||||:::|||
Db        661 GGTGAGGATTTCTTGGAAACCATGGCGTGCGTGCTGCGGCGACCGAGCTACCTCTCGCGG 720

Qy         95 IleGlnValThrLeuLeuCysAsnSerSerAlaGluAlaLeu 108
              :::|||          |||       |||:::      |||:::
Db        721 ATGCAGGGAATCCACCTGCTGAAATCGGCGCTCCCGGCGATG 762

RESULT 3
US-10-004-113-31/c
; Sequence 31, Application US/10004113
; Publication No. US20030194702A1
; GENERAL INFORMATION:
;  APPLICANT: Engelhard, Eric
;  APPLICANT:  Morris, David
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: A-70970/RMS/DCF
;  CURRENT APPLICATION NUMBER: US/10/004,113
;  CURRENT FILING DATE:  2001-10-23
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
;   LENGTH: 26042
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(71)
;   OTHER INFORMATION: "n" at positions 1 through 71 can be any base.
;   NAME/KEY: misc_feature
;   LOCATION: (6720)..(6747)
;   OTHER INFORMATION: "n" at positions 6720 through 6747 can be any base.
;   NAME/KEY: misc_feature
;   LOCATION: (9707)..(10000)
;   OTHER INFORMATION: "n" at positions 9707 through 10000 can be any base.
US-10-004-113-31

Alignment Scores:
Pred. No.:              89.2          Length:        26042
Score:                  78.50         Matches:       25
Percent Similarity:     54.17%        Conservative:  14
Best Local Similarity:  34.72%        Mismatches:    26
Query Match:            14.35%        Indels:        7
```

```
DB:                 16          Gaps:          2

US-10-679-865A-2 (1-108) x US-10-004-113-31 (1-26042)

Qy         30 GluArgAlaSerGlySerSerIleTyrProGluAla--------------LysAlaArg 44
              :::|||   ||||||||||:::   ||||||   |||                  |||
Db      22203 CAAAGGCTCTCGGGCAGCACCTCATACCCAGTGGCTCTTGATCTGACAAACTTAGCAGAA 22144

Qy         45 LeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuProLysGluThrVal 64
                 |||:::   |||   ||||||:::         |||:::   :::   ||||||
Db      22143 TGTTTTCTGCAGTTGCTATACGGAAGTGACAAC------ATGGTGGAAAGGGGGACCGTG 22090

Qy         65 LeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsnValProGlyHis 84
                 :::|||:::::::::   |||   ||||||      :::      |||   |||:::
Db      22089 GGCACTCTTGAGCAGTGGCTGCTGGCATCTTCAGAAACCAGGCTAAGGGTCTTTGGGAAC 22030

Qy         85 ValLeuGlnArgProSerTyrLeuThrArgIleGln 96
                 ||||||   :::|||      |||:::
Db      22029 ACAGCAATCAGGCCTTGTCACTTAAGGCACATTGAA 21994

RESULT 4
US-10-282-122A-26227
; Sequence 26227, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26227
;   LENGTH: 777
;   TYPE: DNA
;   ORGANISM: Mycobacterium bovis
US-10-282-122A-26227

Alignment Scores:
Pred. No.:              0.759          Length:        777
Score:                  77.50          Matches:       34
Percent Similarity:     35.04%         Conservative:  14
Best Local Similarity:  24.82%         Mismatches:    46
Query Match:            14.17%         Indels:        43
DB:                     17             Gaps:          5

US-10-679-865A-2 (1-108) x US-10-282-122A-26227 (1-777)

Qy          5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
              |||      ||||||   |||         |||      :::      :::      |||
Db         49 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTGTAGGCGATCAGGCATCGTTG 108

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
              :::                                |||:::   ||||||        |||
Db        109 TGG-------------------------GAATCCGTGCTGCCCGAGGAGTTGCGGCGG 141

Qy         45 LeuPro--------------LeuGluAlaLeuLeuGlySerAsnLys----------- 55
              ||||||              ::::::|||||||||      :::
Db        142 CTGCCCGAAGAGCTGGCCCGGGTGGATGCGCTGCTCGATGATTCGGCGTTCTTCTGCCCG 201

Qy         56 ----------------------------GluProMetLeuProLysGluThrValLeu 65
                                          |||   :::|||   ||||||   |||
Db        202 TTTGTGCCGTTCTTCGACCCGCGGATGGGTCGGCCGTCCATACCGATGGAGACCTATTTG 261

Qy         66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
                 |||              ||||||   |||||       ::::::   |||   |||
Db        262 CGGTTGATGTTCTTGAAGTTCCGTTACCGGTTGGGCTATGAGTCGCTGTGTCGGGAGGTC 321

Qy         82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                      :::      |||                                  |||
Db        322 ACCGATTCGATCACCTGGCGGCGGTTCTGCCGTATTCCGTTGGAGGGATCGGTGCCGCAC 381

Qy         91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAlaGluAla 107
                 |||   :::::::::|||      |||         |||   |||
Db        382 CCAACCACGTTGATGAAGCTGACCACGCGCTGCGGTGAGGATGCGGTGGCC 432

RESULT 5
US-10-041-018-379
; Sequence 379, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
;  APPLICANT: Matsuda, Seiichi P.T.
;  APPLICANT:  Hart, Elizabeth A.
;  TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
;  FILE REFERENCE: P02080US1/10025547
;  CURRENT APPLICATION NUMBER: US/10/041,018
;  CURRENT FILING DATE:  2002-01-07
;  PRIOR APPLICATION NUMBER: US 60/259880
;  PRIOR FILING DATE: 2001-01-05
;  NUMBER OF SEQ ID NOS: 413
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 379
;   LENGTH: 69350
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-10-041-018-379

Alignment Scores:
Pred. No.:              520            Length:        69350
Score:                  77.50          Matches:       34
Percent Similarity:     35.04%         Conservative:  14
Best Local Similarity:  24.82%         Mismatches:    46
Query Match:            14.17%         Indels:        43
DB:                     18             Gaps:          5

US-10-679-865A-2 (1-108) x US-10-041-018-379 (1-69350)

Qy          5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
              |||      ||||||   |||         |||      :::      :::      |||
Db      57011 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTGTAGGCGATCAGGCATCGTTG 57070

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
              :::                                |||:::   ||||||        |||
Db      57071 TGG-------------------------GAATCCGTGCTGCCCGAGGAGTTGCGGCGG 57103
```

```
Qy          45 LeuPro--------------LeuGluAlaLeuLeuGlySerAsnLys------------ 55
               ||||||               ::::::||||||||     :::
Db       57104 CTGCCCGAAGAGCTGGCCCGGGTGGATGCGCTGCTCGATGATTCGGCGTTCTTCTGCCCG 57163

Qy          56 ----------------------------GluProMetLeuProLysGluThrValLeu 65
                                           |||   :::|||   ||||||   |||
Db       57164 TTTGTGCCGTTCTTCGACCCGCGGATGGGTCGGCCGTCCATACCGATGGAGACCTATTTG 57223

Qy          66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
               |||               ||||||  ||||||     ::::::  |||  |||
Db       57224 CGGTTGATGTTCTTGAAGTTCCGTTACCGGTTGGGCTATGAGTCGCTGTGTCGGGAGGTC 57283

Qy          82 ProGlyHisValLeuGlnArg---------------------------------ProSer 90
                  :::     |||                                       |||
Db       57284 ACCGATTCGATCACCTGGCGGCGGTTCTGCCGTATTCCGTTGGAGGGATCGGTGCCGCAC 57343

Qy          91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAlaGluAla 107
                  |||   :::::::::|||      |||         |||   |||
Db       57344 CCAACCACGTTGATGAAGCTGACCACGCGCTGCGGTGAGGATGCGGTGGCC 57394

RESULT 6
US-10-080-170-649/c
; Sequence 649, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
;  APPLICANT: COLE, S.T.
;  TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
;  TITLE OF INVENTION:  IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
;  TITLE OF INVENTION:  TREATMENT OF MYCOBACTERIOSES
;  FILE REFERENCE: 03495.0218
;  CURRENT APPLICATION NUMBER: US/10/080,170
;  CURRENT FILING DATE:  2002-06-10
;  PRIOR APPLICATION NUMBER: 60/270,123
;  PRIOR FILING DATE: 2001-02-22
;  NUMBER OF SEQ ID NOS: 652
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 649
;   LENGTH: 45191
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-10-080-170-649

Alignment Scores:
Pred. No.:              331          Length:         45191
Score:                  77.00        Matches:        37
Percent Similarity:     34.97%       Conservative:   13
Best Local Similarity:  25.87%       Mismatches:     45
Query Match:            14.08%       Indels:         48
DB:                     15           Gaps:           6

US-10-679-865A-2 (1-108) x US-10-080-170-649 (1-45191)

Qy           5 LeuProIleGluGlyGlnGluValHisPheGluIleLeuGluSerGluCysValLeu 24
               |||      ||||||   |||         |||      |||      :::|||   |||
Db       35763 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTTTAGGCGATCAGTGCTGGTTG 35704

Qy          25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
               :::                                  |||:::   |||:::       |||
Db       35703 TGG--------------------------GAGTCGCTGCTGCCCGACGAGGTGCGCCGA 35671

Qy          45 LeuPro--------------LeuGluAlaLeuLeuGlySer----------------- 53
               ||||||               ::::::|||||||||
Db       35670 CTGCCCGAGGAACTGGCCCGGGTGGACGCATTGTTGGACGATCCGGCGTTCTTCGCCCCG 35611

Qy          54 ----------------------AsnLysGluProMetLeuProLysGluThrValLeu 65
                                       |||      |||   |||      |||
Db       35610 TTCGTGCCGTTCTTCGACCCGCGCAGGGGCCGGCCGTCGACGCCGATGGAGGTCTATCTG 35551

Qy          66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
               |||               ||||||  ||||||     ::::::  |||  |||
Db       35550 CAGTTGATGTTTGTGAAGTTCCGCTACCGGCTGGGCTATGAGTCGCTGTGCCGGGAGGTG 35491

Qy          82 ProGlyHisValLeuGlnArg---------------------------------ProSer 90
                  :::     |||                                       |||
Db       35490 GCTGATTCGATCACCTGACGGCGGTTTTGCCGCATTGCGCTGGACGGGTCGGTGCCGCAT 35431

Qy          91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAla--------------- 105
                  |||   :::::::::|||      |||   |||:::|||
Db       35430 CCGACCACATTGATGAAGCTCACCACGCGTTGCGGATCGGCGGCGATCGACGGGCTCAAC 35371

Qy         106 GluAlaLeu 108
               |||||||||
Db       35370 GAGGCATTG 35362

RESULT 7
US-10-080-170-649/c
; Sequence 649, Application US/10080170
; Publication No. US20040121322A9
; GENERAL INFORMATION:
;  APPLICANT: COLE, S.T.
;  TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
;  TITLE OF INVENTION:  IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
;  TITLE OF INVENTION:  TREATMENT OF MYCOBACTERIOSES
;  FILE REFERENCE: 03495.0218
;  CURRENT APPLICATION NUMBER: US/10/080,170
;  CURRENT FILING DATE:  2002-06-10
;  PRIOR APPLICATION NUMBER: 60/270,123
;  PRIOR FILING DATE: 2001-02-22
;  NUMBER OF SEQ ID NOS: 652
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 649
;   LENGTH: 45191
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-10-080-170-649

Alignment Scores:
Pred. No.:              331          Length:         45191
Score:                  77.00        Matches:        37
Percent Similarity:     34.97%       Conservative:   13
Best Local Similarity:  25.87%       Mismatches:     45
Query Match:            14.08%       Indels:         48
DB:                     19           Gaps:           6

US-10-679-865A-2 (1-108) x US-10-080-170-649 (1-45191)

Qy           5 LeuProIleGluGlyGlnGluValHisPheGluIleLeuGluSerGluCysValLeu 24
               |||      ||||||   |||         |||      |||      :::|||   |||
Db       35763 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTTTAGGCGATCAGTGCTGGTTG 35704

Qy          25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
               :::                                  |||:::   |||:::       |||
Db       35703 TGG--------------------------GAGTCGCTGCTGCCCGACGAGGTGCGCCGA 35671

Qy          45 LeuPro--------------LeuGluAlaLeuLeuGlySer----------------- 53
               ||||||               ::::::|||||||||
Db       35670 CTGCCCGAGGAACTGGCCCGGGTGGACGCATTGTTGGACGATCCGGCGTTCTTCGCCCCG 35611

Qy          54 ----------------------AsnLysGluProMetLeuProLysGluThrValLeu 65
                                       |||      |||   |||      |||
Db       35610 TTCGTGCCGTTCTTCGACCCGCGCAGGGGCCGGCCGTCGACGCCGATGGAGGTCTATCTG 35551

Qy          66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
               |||               ||||||  ||||||     ::::::  |||  |||
Db       35550 CAGTTGATGTTTGTGAAGTTCCGCTACCGGCTGGGCTATGAGTCGCTGTGCCGGGAGGTG 35491

Qy          82 ProGlyHisValLeuGlnArg---------------------------------ProSer 90
                  :::     |||                                       |||
Db       35490 GCTGATTCGATCACCTGACGGCGGTTTTGCCGCATTGCGCTGGACGGGTCGGTGCCGCAT 35431

Qy          91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAla--------------- 105
                  |||   :::::::::|||      |||   |||:::|||
Db       35430 CCGACCACATTGATGAAGCTCACCACGCGTTGCGGATCGGCGGCGATCGACGGGCTCAAC 35371

Qy         106 GluAlaLeu 108
```

```
              |||||||||
Db      35370 GAGGCATTG 35362

RESULT 8
US-10-468-356-649/c
; Sequence 649, Application US/10468356
; Publication No. US20040197896A1
; GENERAL INFORMATION:
;  APPLICANT: COLE, STEWART
;  TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENEOMICS AS A TOOL FOR
;  TITLE OF INVENTION:  IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
;  TITLE OF INVENTION:  TREATMENT OF MYCOBACTERIOSES
;  FILE REFERENCE: 05394.0019
;  CURRENT APPLICATION NUMBER: US/10/468,356
;  CURRENT FILING DATE:  2003-08-19
;  PRIOR APPLICATION NUMBER: 10/080,170
;  PRIOR FILING DATE: 2002-02-22
;  PRIOR APPLICATION NUMBER: 60/270,123
;  PRIOR FILING DATE: 2001-02-22
;  NUMBER OF SEQ ID NOS: 655
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 649
;   LENGTH: 45191
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-10-468-356-649

Alignment Scores:
Pred. No.:                331           Length:        45191
Score:                    77.00         Matches:       37
Percent Similarity:       34.97%        Conservative:  13
Best Local Similarity:    25.87%        Mismatches:    45
Query Match:              14.08%        Indels:        48
DB:                       19            Gaps:          6

US-10-679-865A-2 (1-108) x US-10-468-356-649 (1-45191)

Qy           5 LeuProIleGluGlyGlnGluValHisPheGluIleLeuGluSerGluCysValLeu 24
               |||      ||||||   |||         |||      |||      :::|||   |||
Db       35763 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTTTAGGCGATCAGTGCTGGTTG 35704

Qy          25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
               :::                              |||:::  |||:::       |||
Db       35703 TGG--------------------------GAGTCGCTGCTGCCCGACGAGGTGCGCCGA 35671

Qy          45 LeuPro--------------LeuGluAlaLeuLeuGlySer----------------- 53
               ||||||            ::::::|||||||||
Db       35670 CTGCCCGAGGAACTGGCCCGGGTGGACGCATTGTTGGACGATCCGGCGTTCTTCGCCCCG 35611

Qy          54 ------------------------AsnLysGluProMetLeuProLysGluThrValLeu 65
                                       |||      |||   |||      |||
Db       35610 TTCGTGCCGTTCTTCGACCCGCGCAGGGGCCGGCCGTCGACGCCGATGGAGGTCTATCTG 35551

Qy          66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
                 |||              ||||||   ||||||      ::::::   |||   |||
Db       35550 CAGTTGATGTTTGTGAAGTTCCGCTACCGGCTGGGCTATGAGTCGCTGTGCCGGGAGGTG 35491

Qy          82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                  :::      |||                                     |||
Db       35490 GCTGATTCGATCACCTGACGGCGGTTTTGCCGCATTGCGCTGGACGGGTCGGTGCCGCAT 35431

Qy          91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAla--------------- 105
                  |||   :::::::::|||      |||   |||:::|||
Db       35430 CCGACCACATTGATGAAGCTCACCACGCGTTGCGGATCGGCGGCGATCGACGGGCTCAAC 35371

Qy         106 GluAlaLeu 108
               |||||||||
Db       35370 GAGGCATTG 35362

RESULT 9
US-10-767-701-10785
; Sequence 10785, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 10785
;   LENGTH: 1056
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS39265_1
US-10-767-701-10785

Alignment Scores:
Pred. No.:                1.97          Length:        1056
Score:                    76.00         Matches:       33
Percent Similarity:       46.09%        Conservative:  20
Best Local Similarity:    28.70%        Mismatches:    46
Query Match:              13.89%        Indels:        16
DB:                       19            Gaps:          6

US-10-679-865A-2 (1-108) x US-10-767-701-10785 (1-1056)

Qy           1 MetValThrLeuLeuProIleGluGly---GlnGluValHisPheGluIleLeuGlu 19
               ||||||      |||   :::|||      ||||||      :::
Db       321 ATGGTTGGAGAACTTGAACTGGAAGCTATCCAAGAAGGTCGCTATCTTGCAGATTTCAGA 380

Qy          20 SerGluCysValLeuTyrThrGlyTrpIle---GluArgAlaSerGlySerSerIleTyr 38
                  |||:::   |||:::  ::::::  ::::::  :::  :::   |||
Db       381 TATGGATGCATTGGCTATTCAATGTACTTAGATGACAAGGAAACTCCCACAGGGATACGA 440

Qy          39 ProGluAlaLysAlaArgLeuProLeuGluAla---------LeuLeuGlySerAsnLys 55
                  :::|||:::|||:::|||||||:::   |||         ||||||         |||
Db       441 GATAAGGCACGTGCACAACTACCTGTTTGTGCTGGAGTTGAGCTATTGGCAGATAGAAAA 500

Qy          56 GluProMetLeuProLysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSer 75
                  |||         |||:::      |||      :::   |||      |||
Db       501 GTCCCAACAAATGTTAAGAAAGAAGATCTAAAACGACGTCATTACAAACCAGGGCAA--- 557

Qy          76 AlaMetLysArgAsnValProGlyHisValLeuGlnArgProSerTyrLeuThrArgIle 95
                                  ||||||||||||                :::|||||:::
Db       558 ------------------CCAGGGCACGTG------GGAGATGATTTCTTAACTAAATTT 593

Qy          96 Gln------ValThrLeuLeuCysAsnSerSerAlaGluAlaLeu 108
               |||         |||:::   |||      |||:::   |||
Db       594 CAGAGGAATGCAGGGCTTGTGGCGAATGGTGTGGCTAAGAACCTG 638

RESULT 10
US-09-918-995-32849
; Sequence 32849, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32849
;   LENGTH: 466
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (1)...(466)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-32849

Alignment Scores:
Pred. No.:              1           Length:        466
Score:                  74.50       Matches:       28
Percent Similarity:     47.37%      Conservative:  17
Best Local Similarity:  29.47%      Mismatches:    38
Query Match:            13.62%      Indels:        12
DB:                     10          Gaps:          4

US-10-679-865A-2 (1-108) x US-09-918-995-32849 (1-466)

Qy         18 LeuGluSerGluCysValLeuTyrThrGlyTrpIleGluArgAlaSer-GlySerSerIl 37
Db         12 CTGGCTTCTGAGGCCATTTCCTTCCGCGGCTGG------CGCCACAGCCACCACCAGCAC 65

Qy         37 eTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySer---------As 54
Db         66 GTCCTCCACGGTCACCGTCACGGCCCCGGCCCCCGCCGCCACAGGAAGCCCAGTGAAAAA 125

Qy         54 nLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySe 74
Db        126 GCAGAGGCCGCTTTTACCGAAGGAGACTGCCCCGGCCGTGCAGCGGGTCGTGTGGAACTC 185

Qy         74 rSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgProSerTyrLeuThrAr 94
Db        186 ATCAACTGTCCAGCAGAAG---------GAGATCACACAGAGCCCATCCACGTCCACC-- 234

Qy         94 gIleGlnValThrLeuLeuCysAsnSerSerAlaGluAlaLeu 108
Db        235 -------ATCACCCTGGTGACCAGCACACAGTCATCGCCCCTG 270

RESULT 11
US-10-719-993-7051/c
; Sequence 7051, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7051
;   LENGTH: 52128
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(52128)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-7051

Alignment Scores:
Pred. No.:              952         Length:        52128
Score:                  74.50       Matches:       21
Percent Similarity:     47.30%      Conservative:  14
Best Local Similarity:  28.38%      Mismatches:    26
Query Match:            13.62%      Indels:        13
DB:                     20          Gaps:          2

US-10-679-865A-2 (1-108) x US-10-719-993-7051 (1-52128)

Qy         10 GlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu--------------- 24
Db       6246 CAATGCACACACACATATAAGTATATATGGAATGAATGCATCTTAGCTTTGCTGCTTATT 6187

Qy         25 ---------------------TyrThrGlyTrpIleGluArgAlaSerGlySerSerIle 37
```

```
Db       6186 AGATATGGTCCTTGAAAAAATTACTCGGGCAARTTTGTGAAAGTGCCTGGCACAAGGTGT 6127

Qy         38 TyrProGluAlaLysAlaArgLeu---ProLeuGluAlaLeuLeuGlySerAsnLysGlu 56
Db       6126 TTCGCAGAAGCAAAAAGCCGTCTTTTTCCAGTCACAGCTATTCAAGGTATAGTAAGTGAA 6067

Qy         57 ProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
Db       6066 TCAAATGAGCCCCTTACCTCCCATCTACATATTCCCCGATAT 6025

RESULT 12
US-10-424-599-28834/c
; Sequence 28834, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 28834
;   LENGTH: 361
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_126038C.1
US-10-424-599-28834

Alignment Scores:
Pred. No.:              0.817       Length:        361
Score:                  74.00       Matches:       19
Percent Similarity:     38.89%      Conservative:  9
Best Local Similarity:  26.39%      Mismatches:    16
Query Match:            13.53%      Indels:        28
DB:                     18          Gaps:          2

US-10-679-865A-2 (1-108) x US-10-424-599-28834 (1-361)

Qy          2 ValThrLeuLeuProIleGluGlyGlnGluVal--------------------------- 12
Db        230 GTAACGCTGGTACCAATCGAAGGAACCACCCTCTTGGATGTAGCGCGCGTCCCAATACCC 171

Qy         13 ---------------------------------------------HisPhePheGluIle 17
Db        170 AGCCTCGCCGTAGTTGTACGTGTTGCAGCTTGACACGTCCCTAAACATCTTTTCAATCTT 111

Qy         18 LeuGluSerGluCysValLeuTyrThrGlyTrp-----------IleGluArgAlaSer 33
Db        110 ATCTTATCACGGTGTGTAATCGATTATGGCTGGTGTGTGCTTCGAGTGGAATACGCCGGC 51

Qy         34 GlySerSerIleTyrProGluAlaLysAlaArgLeu 45
Db         50 GACTCACAACACCACCCACGCGTCCGCGGAAGAGTG 15

RESULT 13
US-10-108-260A-2017/c
; Sequence 2017, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2017
;   LENGTH: 1682
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-2017

Alignment Scores:
Pred. No.:                7.66          Length:        1682
Score:                    74.00         Matches:       22
Percent Similarity:       50.00%        Conservative:  12
Best Local Similarity:    32.35%        Mismatches:    34
Query Match:              13.53%        Indels:        0
DB:                       17            Gaps:          0

US-10-679-865A-2 (1-108) x US-10-108-260A-2017 (1-1682)

Qy          30 GluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAla 49
               ||||||||||:::  ||||||    |||:::|||  :::
Db         707 GAGCGGGCGGCGGAGAGCTCCCGGGCGCCCGACGCCGGGTCCGGGAGCTCAGCCCGCAGG 648

Qy          50 LeuLeuGlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArg 69
               ||||||          |||  |||:::||||||       :::  :::      :::
Db         647 TTGCTGGAGATGCAGAAGGCCCCTCTTTTGCCCCTAGCCCTCCTCGGGGCTTCTCTCCAG 588

Qy          70 TyrAsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeuGlnArgPro 89
                  :::|||   |||      ::::::  :::     ||||||         ||||||
Db         587 TAAAGCCTGAGGAGCCGGCGTCTGGAGTGCAGTGCGAATGGTCACACTTGTGACCGCCCC 528

Qy          90 SerTyrLeuThrArgIleGlnVal 97
               |||      :::|||      |||
Db         527 TCTGCAGCCTCGCGCTGCTTAGTC 504

RESULT 14
US-10-027-632-98121/c
; Sequence 98121, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 98121
;   LENGTH: 1926
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-98121

Alignment Scores:
Pred. No.:                9.32          Length:        1926
Score:                    74.00         Matches:       31
Percent Similarity:       41.51%        Conservative:  13
Best Local Similarity:    29.25%        Mismatches:    38
Query Match:              13.53%        Indels:        24
DB:                       13            Gaps:          5

US-10-679-865A-2 (1-108) x US-10-027-632-98121 (1-1926)

Qy          12 ValHisPhePheGluIleLeuGlu----------------------------SerGlu 21
               |||||||||       :::   :::                           |||
Db         460 GTGCACTTCTTACACCTGACGAAGCAGCACCAGTGGAACTTGCAGTGGCAGCGCTCCACC 401

Qy          22 CysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAla 41
               |||   |||                 |||           |||     |||
Db         400 TGCACGCTCTTGAACTGGTTGTAGCCACGCCCGCAGCACATGAGCTCACAGCCATCC--- 344

Qy          42 LysAlaArgLeuProLeuGluAlaLeuGlySerAsnLysGluProMetLeuProLys 61
                  :::|||   |||  ||||||       :::   |||  :::|||:::
Db         343 ---------ATGCCCTCCGAGGTCTTGTTG------CAGAGGCGGCCCTGCGTGCCCAGG 299

Qy          62 GluThrValLeuSerLeuLysArgTyr------AsnLeuGlySerSerAlaMetLysArg 79
               |||   ||||||||||||:::|||          |||||||||:::     :::
Db         298 GAGCCCGTGCTCTCGTTGCGCAGGCAGTAGTCGGGGCTGGGGTCCACATAGACCAGGTCC 239

Qy          80 AsnValProGlyHisValLeuGln-----------------ArgProSerTyrLeuThr 93
               :::      |||  |||  :::                  ||||||     :::|||
Db         238 TCCGGGGTGGGCTGGGTGAAGCGGCTGTTGACCAGCTCCAGCCGGCCCTTGCGGGTGACG 179

Qy          94 ArgIleGlnValThrLeu 99
               |||:::         |||
Db         178 CGCATGGCGGCCGCGCTG 161

RESULT 15
US-10-027-632-98122/c
; Sequence 98122, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 98122
;   LENGTH: 1926
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-98122

Alignment Scores:
Pred. No.:                9.32          Length:        1926
Score:                    74.00         Matches:       31
Percent Similarity:       41.51%        Conservative:  13
Best Local Similarity:    29.25%        Mismatches:    38
Query Match:              13.53%        Indels:        24
DB:                       13            Gaps:          5

US-10-679-865A-2 (1-108) x US-10-027-632-98122 (1-1926)

Qy          12 ValHisPhePheGluIleLeuGlu----------------------------SerGlu 21
               |||||||||       :::   :::                           |||
Db         460 GTGCACTTCTTACACCTGACGAAGCAGCACCAGTGGAACTTGCAGTGGCAGCGCTCCACC 401

Qy          22 CysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAla 41
               |||   |||                 |||           |||     |||
```

```
Db        400 TGCACGCTCTTGAACTGGTTGTAGCCACGCCCGCAGCACATGAGCTCACAGCCATCC--- 344

Qy         42 LysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuProLys 61
                  :::|||   |||   ||||||         :::   |||   :::|||:::
Db        343 ---------ATGCCCTCCGAGGTCTTGTTG------CAGAGGCGGCCCTGCGTGCCCAGG 299

Qy         62 GluThrValLeuSerLeuLysArgTyr------AsnLeuGlySerSerAlaMetLysArg 79
              |||   ||||||||||||:::|||           |||||||||:::      :::
Db        298 GAGCCCGTGCTCTCGTTGCGCAGGCAGTAGTCGGGGCTGGGGTCCACATAGACCAGGTCC 239

Qy         80 AsnValProGlyHisValLeuGln-----------------ArgProSerTyrLeuThr 93
              :::      |||   |||   :::                 ||||||      :::|||
Db        238 TCCGGGGTGGGCTGGGTGAAGCGGCTGTTGACCAGCTCCAGCCGGCCCTTGCGGGTGACG 179

Qy         94 ArgIleGlnValThrLeu 99
              |||:::         |||
Db        178 CGCATGGCGGCCGCGCTG 161
```

Search completed: August 31, 2005, 23:15:15
Job time : 588 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 31, 2005, 19:53:22 ; Search time 594 Seconds
(without alignments)
297.505 Million cell updates/sec

Title: US-10-679-865A-2
Perfect score: 547
Sequence: 1 MVTLLPIEGQEVHFFEILES.....PSYLTRIQVTLLCNSSAEAL 108

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US10679865/runat_30082005_150802_15923/app_query.fasta_1.263
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10679865_@CGN_1_1_69_@runat_30082005_150802_15923 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	80.5	14.7	50725	4	US-09-902-540-1271	Sequence 1271, Ap
	2	77.5	14.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	3	77.5	14.2	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	4	77	14.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	5	77	14.1	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	6	74	13.5	2057	3	US-09-082-089-4	Sequence 4, Appli
c	7	74	13.5	2190	3	US-09-082-089-1	Sequence 1, Appli
	8	73	13.3	601	4	US-09-949-016-22538	Sequence 22538, A
	9	73	13.3	601	4	US-09-949-016-41489	Sequence 41489, A
c	10	72.5	13.3	11272	4	US-09-341-461-1	Sequence 1, Appli
	11	72	13.2	7647	4	US-09-566-921-75	Sequence 75, Appl
c	12	71.5	13.1	1302	4	US-09-252-991A-3473	Sequence 3473, Ap
	13	71.5	13.1	1332	4	US-09-252-991A-3464	Sequence 3464, Ap
c	14	71.5	13.1	1722	4	US-09-252-991A-3485	Sequence 3485, Ap
	15	71.5	13.1	3198	4	US-09-902-540-2522	Sequence 2522, Ap
c	16	71.5	13.1	17503	4	US-09-902-540-1114	Sequence 1114, Ap
	17	71.5	13.1	25198	4	US-09-949-016-16575	Sequence 16575, A
	18	71	13.0	417	4	US-09-252-991A-3455	Sequence 3455, Ap
c	19	71	13.0	1026	4	US-09-252-991A-3491	Sequence 3491, Ap
	20	71	13.0	10432	4	US-09-919-172-97	Sequence 97, Appl
	21	71	13.0	10432	4	US-09-976-594-21	Sequence 21, Appl
	22	71	13.0	10432	4	US-09-919-039-20	Sequence 20, Appl
c	23	71	13.0	69834	4	US-09-949-016-11959	Sequence 11959, A
c	24	71	13.0	69834	4	US-09-949-016-12925	Sequence 12925, A
	25	70.5	12.9	1584	4	US-09-328-352-3607	Sequence 3607, Ap
	26	70	12.8	57978	4	US-09-949-016-16667	Sequence 16667, A
c	27	70	12.8	240157	4	US-09-949-016-16264	Sequence 16264, A
	28	69.5	12.7	2860	3	US-09-221-017B-871	Sequence 871, App
	29	69.5	12.7	152481	4	US-09-949-016-12521	Sequence 12521, A
	30	69.5	12.7	152798	4	US-09-949-016-12775	Sequence 12775, A
	31	69.5	12.7	152822	4	US-09-949-016-17518	Sequence 17518, A
	32	69.5	12.7	152822	4	US-09-949-016-17519	Sequence 17519, A
	33	69	12.6	687	4	US-09-252-991A-8427	Sequence 8427, Ap
c	34	69	12.6	2496	1	US-08-073-384C-2	Sequence 2, Appli
c	35	69	12.6	2496	1	US-08-254-359A-2	Sequence 2, Appli
c	36	69	12.6	2496	1	US-08-483-043-2	Sequence 2, Appli
c	37	69	12.6	2496	1	US-08-481-238-2	Sequence 2, Appli
c	38	69	12.6	2496	2	US-08-471-066B-2	Sequence 2, Appli
c	39	69	12.6	2496	2	US-08-484-956-2	Sequence 2, Appli
c	40	69	12.6	2496	2	US-08-757-653-2	Sequence 2, Appli
c	41	69	12.6	2496	2	US-08-599-491-2	Sequence 2, Appli
c	42	69	12.6	2496	2	US-08-756-386-2	Sequence 2, Appli
c	43	69	12.6	2496	2	US-08-823-516-2	Sequence 2, Appli
c	44	69	12.6	2496	3	US-08-682-853A-2	Sequence 2, Appli
c	45	69	12.6	2496	3	US-08-759-038-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-902-540-1271/c
; Sequence 1271, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1271
; LENGTH: 50725
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1271

Alignment Scores:

Pred. No.:	113	Length:	50725
Score:	80.50	Matches:	29
Percent Similarity:	41.67%	Conservative:	6
Best Local Similarity:	34.52%	Mismatches:	26
Query Match:	14.72%	Indels:	23
DB:	4	Gaps:	4

US-10-679-865A-2 (1-108) x US-09-902-540-1271 (1-50725)

```
Qy          5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
              ||||||:::|||   |||   |||   |||||||||:::||||||   |||
Db      21564 CTGCCGCTGGAGGCGCAGCACGTGAGGTTCTTCGAGGTTCTCGAGGATGAAGCCACCCGG 21505
```

```
Qy          25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIle---------- 37
               ||||||   |||:::   |||     |||
Db       21504 AACACGGGCCGTATCGACGGAGTGAGCCCCTGGTGCATCCCGAGCATCGACTGCACCGTC 21445

Qy          38 ----------TyrProGluAlaLys---AlaArgLeuPro 46
                        ||||||        ||||||
Db       21444 TGTGGCGGCTACGGCGGCCTCGGAGAGGCCTACCCCACTGTGGACCTCGCGCGTTTCGCA 21385

Qy          47 LeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuProLysGluThrValLeuSer 66
               |||||||||      :::||||||      :::      ||||||
Db       21384 CACCGTGCCCTGCTG---------GAAGAACCGCGG------AGCCAACAAGCCCTGTCA 21340

Qy          67 LeuLysArgTyr 70
               |||:::   |||
Db       21339 CTGGAGGCGTAC 21328

RESULT 2
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.:              2.58e+05      Length:         4403765
Score:                  77.50         Matches:        34
Percent Similarity:     35.04%        Conservative:   14
Best Local Similarity:  24.82%        Mismatches:     46
Query Match:            14.17%        Indels:         43
DB:                     3             Gaps:           5

US-10-679-865A-2 (1-108) x US-09-103-840A-2 (1-4403765)

Qy           5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
               |||     ||||||  |||       |||     :::     :::       |||
Db     3790883 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTGTAGGCGATCAGGCATCGTTG 3790942

Qy          25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
               :::                             |||:::   ||||||        |||
Db     3790943 TGG-------------------------GAATCCGTGCTGCCCGAGGAGTTGCGGCGG 3790975

Qy          45 LeuPro--------------LeuGluAlaLeuLeuGlySerAsnLys---------- 55
               ||||||              ::::::|||||||||     :::
Db     3790976 CTGCCCGAAGAGCTGGCCCGGGTGGATGCGCTGCTCGATGATTCGGCGTTCTTCTGCCCG 3791035

Qy          56 ----------------------------GluProMetLeuProLysGluThrValLeu 65
                                           |||   :::|||   ||||||   |||
Db     3791036 TTTGTGCCGTTCTTCGACCCGCGGATGGGTCGGCCGTCCATACCGATGGAGACCTATTTG 3791095

Qy          66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
               |||               ||||||  ||||||      ::::::  |||   |||
Db     3791096 CGGTTGATGTTCTTGAAGTTCCGTTACCGGTTGGGCTATGAGTCGCTGTGTCGGGAGGTC 3791155

Qy          82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                        :::      |||                                  |||
Db     3791156 ACCGATTCGATCACCTGGCGGCGGTTCTGCCGTATTCCGTTGGAGGGATCGGTGCCGCAC 3791215

Qy          91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAlaGluAla 107
               |||   :::::::::|||      |||          |||   |||
Db     3791216 CCAACCACGTTGATGAAGCTGACCACGCGCTGCGGTGAGGATGCGGTGGCC 3791266

RESULT 3
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.:              2.59e+05      Length:         4411529
Score:                  77.50         Matches:        34
Percent Similarity:     35.04%        Conservative:   14
Best Local Similarity:  24.82%        Mismatches:     46
Query Match:            14.17%        Indels:         43
DB:                     3             Gaps:           5

US-10-679-865A-2 (1-108) x US-09-103-840A-1 (1-4411529)

Qy           5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
               |||     ||||||  |||       |||     :::     :::       |||
Db     3800061 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTGTAGGCGATCAGGCATCGTTG 3800120

Qy          25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
               :::                             |||:::   ||||||        |||
Db     3800121 TGG-------------------------GAATCCGTGCTGCCCGAGGAGTTGCGGCGG 3800153

Qy          45 LeuPro--------------LeuGluAlaLeuLeuGlySerAsnLys---------- 55
               ||||||              ::::::|||||||||     :::
Db     3800154 CTGCCCGAAGAGCTGGCCCGGGTGGATGCGCTGCTCGATGATTCGGCGTTCTTCTGCCCG 3800213

Qy          56 ----------------------------GluProMetLeuProLysGluThrValLeu 65
                                           |||   :::|||   ||||||   |||
Db     3800214 TTTGTGCCGTTCTTCGACCCGCGGATGGGTCGGCCGTCCATACCGATGGAGACCTATTTG 3800273

Qy          66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
               |||               ||||||  ||||||      ::::::  |||   |||
Db     3800274 CGGTTGATGTTCTTGAAGTTCCGTTACCGGTTGGGCTATGAGTCGCTGTGTCGGGAGGTC 3800333

Qy          82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                        :::      |||                                  |||
Db     3800334 ACCGATTCGATCACCTGGCGGCGGTTCTGCCGTATTCCGTTGGAGGGATCGGTGCCGCAC 3800393

Qy          91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAlaGluAla 107
               |||   :::::::::|||      |||          |||   |||
Db     3800394 CCAACCACGTTGATGAAGCTGACCACGCGCTGCGGTGAGGATGCGGTGGCC 3800444

RESULT 4
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
```

; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:

Pred. No.:	3e+05	Length:	4403765
Score:	77.00	Matches:	37
Percent Similarity:	34.97%	Conservative:	13
Best Local Similarity:	25.87%	Mismatches:	45
Query Match:	14.08%	Indels:	48
DB:	3	Gaps:	6

US-10-679-865A-2 (1-108) x US-09-103-840A-2 (1-4403765)

```
Qy          5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
              |||      ||||||  |||        |||      |||      :::|||  |||
Db    1160408 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTTTAGGCGATCAGTGCTGGTTG 1160349

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
              :::                          |||:::  |||:::        |||
Db    1160348 TGG-------------------------GAGTCGCTGCTGCCCGACGAGGTGCGCCGA 1160316

Qy         45 LeuPro--------------LeuGluAlaLeuLeuGlySer----------------- 53
              ||||||              ::::::|||||||||
Db    1160315 CTGCCCGAGGAACTGGCCCGGGTGGACGCATTGTTGGACGATCCGGCGTTCTTCGCCCCG 1160256

Qy         54 ----------------------AsnLysGluProMetLeuProLysGluThrValLeu 65
                                     |||      |||  |||      |||
Db    1160255 TTCGTGCCGTTCTTCGACCCGCGCAGGGGCCGGCCGTCGACGCCGATGGAGGTCTATCTG 1160196

Qy         66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
                |||              ||||||  ||||||      :::::  |||  |||
Db    1160195 CAGTTGATGTTTGTGAAGTTCCGCTACCGGCTGGGCTATGAGTCGCTGTGCCGGGAGGTG 1160136

Qy         82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                  :::      |||                                    |||
Db    1160135 GCTGATTCGATCACCTGACGGCGGTTTTGCCGCATTGCGCTGGACGGGTCGGTGCCGCAT 1160076

Qy         91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAla--------------- 105
                 |||   :::::::::|||      |||   |||:::|||
Db    1160075 CCGACCACATTGATGAAGCTCACCACGCGTTGCGGATCGGCGGCGATCGACGGGCTCAAC 1160016

Qy        106 GluAlaLeu 108
              |||||||||
Db    1160015 GAGGCATTG 1160007
```

RESULT 5
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:

Pred. No.:	3.01e+05	Length:	4411529
Score:	77.00	Matches:	37
Percent Similarity:	34.97%	Conservative:	13
Best Local Similarity:	25.87%	Mismatches:	45
Query Match:	14.08%	Indels:	48
DB:	3	Gaps:	6

US-10-679-865A-2 (1-108) x US-09-103-840A-1 (1-4411529)

```
Qy          5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
              |||      ||||||  |||        |||      |||      :::|||  |||
Db    1160383 CTTGTTGCGGAAGGGATTGAGGCCATCGTGTTTCGTACTTTAGGCGATCAGTGCTGGTTG 1160324

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
              :::                          |||:::  |||:::        |||
Db    1160323 TGG-------------------------GAGTCGCTGCTGCCCGACGAGGTGCGCCGA 1160291

Qy         45 LeuPro--------------LeuGluAlaLeuLeuGlySer----------------- 53
              ||||||              ::::::|||||||||
Db    1160290 CTGCCCGAGGAACTGGCCCGGGTGGACGCATTGTTGGACGATCCGGCGTTCTTCGCCCCG 1160231

Qy         54 ----------------------AsnLysGluProMetLeuProLysGluThrValLeu 65
                                     |||      |||  |||      |||
Db    1160230 TTCGTGCCGTTCTTCGACCCGCGCAGGGGCCGGCCGTCGACGCCGATGGAGGTCTATCTG 1160171

Qy         66 SerLeuLys-----------ArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
                |||              ||||||  ||||||      :::::  |||  |||
Db    1160170 CAGTTGATGTTTGTGAAGTTCCGCTACCGGCTGGGCTATGAGTCGCTGTGCCGGGAGGTG 1160111

Qy         82 ProGlyHisValLeuGlnArg--------------------------------ProSer 90
                  :::      |||                                    |||
Db    1160110 GCTGATTCGATCACCTGACGGCGGTTTTGCCGCATTGCGCTGGACGGGTCGGTGCCGCAT 1160051

Qy         91 TyrLeuThrArgIleGlnValThrLeuLeuCysAsnSerSerAla--------------- 105
                 |||   :::::::::|||      |||   |||:::|||
Db    1160050 CCGACCACATTGATGAAGCTCACCACGCGTTGCGGATCGGCGGCGATCGACGGGCTCAAC 1159991

Qy        106 GluAlaLeu 108
              |||||||||
Db    1159990 GAGGCATTG 1159982
```

RESULT 6
US-09-082-089-4/c
; Sequence 4, Application US/09082089
; Patent No. 6100060
; GENERAL INFORMATION:
; APPLICANT: BARNES, MICHAEL
; APPLICANT: TESTA, TANIA
; APPLICANT: KELSELL, DAVID
; TITLE OF INVENTION: No. 6100060el Compounds
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:

```
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/082,089
;      FILING DATE:  20-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9710737.9
;      FILING DATE:  23-MAY-1997
;      APPLICATION NUMBER:  GB 9803981.1
;      FILING DATE:  25-FEB-1998
;      APPLICATION NUMBER:  GB 9804007.4
;      FILING DATE:  25-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F.
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-30166
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2057 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-082-089-4

Alignment Scores:
Pred. No.:              6.77          Length:         2057
Score:                  74.00         Matches:        31
Percent Similarity:     41.51%        Conservative:   13
Best Local Similarity:  29.25%        Mismatches:     38
Query Match:            13.53%        Indels:         24
DB:                     3             Gaps:           5

US-10-679-865A-2 (1-108) x US-09-082-089-4 (1-2057)

Qy         12 ValHisPhePheGluIleLeuGlu----------------------------SerGlu 21
              |||||||||      :::   :::                            |||
Db       1061 GTGCACTTCTTACACCTGACGAAGCAGCACCAGTGGAACTTGCARTGGCAGCGCTCCACC 1002

Qy         22 CysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAla 41
              |||   |||              |||          |||      |||
Db       1001 TGCACGCTCTTGAACTGGTTGTARCCACGCCCGCAGCACATGAGCTCACAGCCATCC--- 945

Qy         42 LysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuProLys 61
                    :::|||   |||   ||||||         :::   |||   :::|||:::
Db        944 ---------ATGCCCTCCGAGGTCTTGTTG------CAGAGGCGGCCCTGCGTGCCCAGG 900

Qy         62 GluThrValLeuSerLeuLysArgTyr------AsnLeuGlySerSerAlaMetLysArg 79
              |||   |||||||||||:::|||           |||||||:::      :::
Db        899 GAGCCCGTGCTCTCGTTGCGCAGGCAGTAGTCGGGGCTGGGGTCCACATAGACCAGGTCC 840

Qy         80 AsnValProGlyHisValLeuGln-----------------ArgProSerTyrLeuThr 93
              :::     |||   |||   :::                 ||||||      :::|||
Db        839 TCCGGGGTGGGCTGGGTGAAGCGGCTGTTGACCAGCTCCAGCCGGCCCTTGCGGGTGACG 780

Qy         94 ArgIleGlnValThrLeu 99
              |||:::        |||
Db        779 CGCATGGCGGCCGCGCTG 762

RESULT 7
US-09-082-089-1/c
; Sequence 1, Application US/09082089
; Patent No. 6100060
;  GENERAL INFORMATION:
;    APPLICANT:  BARNES, MICHAEL
;    APPLICANT:  TESTA, TANIA
;    APPLICANT:  KELSELL, DAVID
;    TITLE OF INVENTION:  No. 6100060el Compounds
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/082,089
;      FILING DATE:  20-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9710737.9
;      FILING DATE:  23-MAY-1997
;      APPLICATION NUMBER:  GB 9803981.1
;      FILING DATE:  25-FEB-1998
;      APPLICATION NUMBER:  GB 9804007.4
;      FILING DATE:  25-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F.
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-30166
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2190 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-082-089-1

Alignment Scores:
Pred. No.:              7.46          Length:         2190
Score:                  74.00         Matches:        31
Percent Similarity:     41.51%        Conservative:   13
Best Local Similarity:  29.25%        Mismatches:     38
Query Match:            13.53%        Indels:         24
DB:                     3             Gaps:           5

US-10-679-865A-2 (1-108) x US-09-082-089-1 (1-2190)

Qy         12 ValHisPhePheGluIleLeuGlu----------------------------SerGlu 21
              |||||||||      :::   :::                            |||
Db       1194 GTGCACTTCTTACACCTGACGAAGCAGCACCAGTGGAACTTGCAGTGGCAGCGCTCCACC 1135

Qy         22 CysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAla 41
              |||   |||              |||          |||      |||
Db       1134 TGCACGCTCTTGAACTGGTTGTAGCCACGCCCGCAGCACATGAGCTCACAGCCATCC--- 1078

Qy         42 LysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuProLys 61
                    :::|||   |||   ||||||         :::   |||   :::|||:::
Db       1077 ---------ATGCCCTCCGAGGTCTTGTTG------CAGAGGCGGCCCTGCGTGCCCAGG 1033

Qy         62 GluThrValLeuSerLeuLysArgTyr------AsnLeuGlySerSerAlaMetLysArg 79
              |||   |||||||||||:::|||           |||||||:::      :::
Db       1032 GAGCCCGTGCTCTCGTTGCGCAGGCAGTAGTCGGGGCTGGGGTCCACATAGACCAGGTCC 973

Qy         80 AsnValProGlyHisValLeuGln-----------------ArgProSerTyrLeuThr 93
              :::     |||   |||   :::                 ||||||      :::|||
Db        972 TCCGGGGTGGGCTGGGTGAAGCGGCTGTTGACCAGCTCCAGCCGGCCCTTGCGGGTGACG 913
```

```
Qy          94 ArgIleGlnValThrLeu 99
               |||:::          |||
Db         912 CGCATGGCGGCCGCGCTG 895

RESULT 8
US-09-949-016-22538
; Sequence 22538, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22538
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-22538

Alignment Scores:
Pred. No.:              1.42          Length:         601
Score:                  73.00         Matches:        19
Percent Similarity:     41.54%        Conservative:   8
Best Local Similarity:  29.23%        Mismatches:     38
Query Match:            13.35%        Indels:         0
DB:                     4             Gaps:           0

US-10-679-865A-2 (1-108) x US-09-949-016-22538 (1-601)

Qy          27 GlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArgLeuPro 46
               ||||||   |||      :::         :::   ||||||||||||         |||
Db         163 GGGTGGAGTGAAGGTCGGAACCACCAAAGGGTGGACCCAGAAGCTAAGCCGCCAACCCCA 222

Qy          47 LeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuProLysGluThrValLeuSer 66
                           |||         :::|||      ||||||         :::|||
Db         223 TCTCCACTGCCAGAGGGACACATGGACCAGCCACAGGCCCCCAAGACGCAGCAGGTGAGC 282

Qy          67 LeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeu 86
                     :::   |||         |||      |||:::
Db         283 CCAGCTGCCCATGGCCTCSTCAGGAGAGCCTGCAGCCGCCACCCAGCGGCAAGCTGTGCA 342

Qy          87 GlnArgProSerTyr 91
               |||||||||::::::
Db         343 CAGAGGCCGACCCAC 357

RESULT 9
US-09-949-016-41489
; Sequence 41489, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 41489
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-41489

Alignment Scores:
Pred. No.:              1.42          Length:         601
Score:                  73.00         Matches:        19
Percent Similarity:     41.54%        Conservative:   8
Best Local Similarity:  29.23%        Mismatches:     38
Query Match:            13.35%        Indels:         0
DB:                     4             Gaps:           0

US-10-679-865A-2 (1-108) x US-09-949-016-41489 (1-601)

Qy          27 GlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArgLeuPro 46
               ||||||   |||      :::         :::   ||||||||||||         |||
Db         163 GGGTGGAGTGAAGGTCGGAACCACCAAAGGGTGGACCCAGAAGCTAAGCCGCCAACCCCA 222

Qy          47 LeuGluAlaLeuLeuGlySerAsnLysGluProMetLeuProLysGluThrValLeuSer 66
                           |||         :::|||      ||||||         :::|||
Db         223 TCTCCACTGCCAGAGGGACACATGGACCAGCCACAGGCCCCCAAGACGCAGCAGGTGAGC 282

Qy          67 LeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsnValProGlyHisValLeu 86
                     :::   |||         |||      |||:::
Db         283 CCAGCTGCCCATGGCCTCSTCAGGAGAGCCTGCAGCCGCCACCCAGCGGCAAGCTGTGCA 342

Qy          87 GlnArgProSerTyr 91
               |||||||||::::::
Db         343 CAGAGGCCGACCCAC 357

RESULT 10
US-09-341-461-1/c
; Sequence 1, Application US/09341461
; Patent No. 6586389
; GENERAL INFORMATION:
;  APPLICANT:  Hammond, Timothy G.
;  APPLICANT:  Verroust, Pierre J.
;  TITLE OF INVENTION:  Cubilin Protein, DNA Sequences Encoding Cubilin
;  TITLE OF INVENTION:  and Uses Thereof
;  FILE REFERENCE:  D6148
;  CURRENT APPLICATION NUMBER: US/09/341,461
;  CURRENT FILING DATE:  2000-07-20
;  PRIOR APPLICATION NUMBER:  PCT/US99/01259
;  PRIOR FILING DATE:  1999-01-21
;  NUMBER OF SEQ ID NOS:  40
; SEQ ID NO 1
;   LENGTH: 11272
;   TYPE: DNA
;   ORGANISM: rat
;   FEATURE:
;   OTHER INFORMATION: nucleic acid sequence of rat cubilin
US-09-341-461-1

Alignment Scores:
Pred. No.:              148           Length:         11272
Score:                  72.50         Matches:        26
Percent Similarity:     44.19%        Conservative:   12
Best Local Similarity:  30.23%        Mismatches:     25
Query Match:            13.25%        Indels:         23
DB:                     4             Gaps:           4

US-10-679-865A-2 (1-108) x US-09-341-461-1 (1-11272)

Qy          13 HisPhePheGluIleLeuGluSerGluCysValLeuTyrThrGlyTrpIleGluArgAla 32
               ||||||      |||   |||      |||                          |||
Db        1156 CATTTCCTGTATATCCCGGAGGACAGGTGC----------------------AGACAG 1121

Qy          33 SerGlySerSerIleTyrProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeu--- 51
               :::      ::::::      |||   ::::::|||||||||||:::      |||
Db        1120 GCAAGAAGGACCCTAGAACAGGAGATGAGGAGCAGGTTGCCTCTGGATGGCAGCCTCCAT 1061
```

```
Qy          52 ---GlySerAsnLysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyr 70
               |||||||||:::           |||:::|||     ||||||
Db        1060 TGTGGATCGAACAGA---------TGTCCACGGGAGTGCACACTCTCCCGTCACCACTGA 1010

Qy          71 AsnLeuGlySerSerAlaMetLysArgAsnVal-------------------------- 81
               ||||||     |||:::   :::|||   :::
Db        1009 AACCTGCTGGACAGTTCCCACAGGAGAATGATCCAGGTGTGTTCAGGCATGGGACCAGTG 950

Qy          82 ---ProGlyHisValLeu 86
                  ||||||:::   |||
Db         949 GCGCCTGGGAACAGCCTC 932

RESULT 11
US-09-566-921-75
; Sequence 75, Application US/09566921
; Patent No. 6682888
; GENERAL INFORMATION:
;  APPLICANT:  Loring, Jeanne F.
;  APPLICANT:  Tingley, Debora W.
;  APPLICANT:  Edwards, Carla M.
;  TITLE OF INVENTION:  GENES EXPRESSED IN ALZHEIMER'S DISEASE
;  FILE REFERENCE:  PA-0024 US
;  CURRENT APPLICATION NUMBER: US/09/566,921
;  CURRENT FILING DATE:  2000-05-05
;  NUMBER OF SEQ ID NOS:  138
;  SOFTWARE:  PERL Program
; SEQ ID NO 75
;   LENGTH: 7647
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6682888 334247.2
;   NAME/KEY: unsure
;   LOCATION: 4641
;   OTHER INFORMATION: a, t, c, g, or other
US-09-566-921-75

Alignment Scores:
Pred. No.:              95.9          Length:        7647
Score:                  72.00         Matches:       33
Percent Similarity:     37.50%        Conservative:  18
Best Local Similarity:  24.26%        Mismatches:    49
Query Match:            13.16%        Indels:        36
DB:                     4             Gaps:          5

US-10-679-865A-2 (1-108) x US-09-566-921-75 (1-7647)

Qy           6 ProIleGluGlyGlnGluValHisPhePheGlu-----------------------Ile 17
               |||:::||||||:::      |||   :::|||
Db        1712 CCCCTGGAGGGCAGGCGGTATCATATCTATGAAAATGGCACATTGCAGATCAACAGAACC 1771

Qy          18 LeuGluSerGluCysValLeuTyrThrGlyTrpIleGluArgAlaSerGlySerSerIle 37
                  |||   :::         |||:::   |||:::|||   |||   |||   :::
Db        1772 ACCGAAGAAGATGCTGGGTCTTACTCATGTTGGGTAGAAAATGCTATAGGAAAAACT--- 1828

Qy          38 TyrProGluAlaLysAlaArgLeuProLeuGlu---------AlaLeuLeuGlySerAsn 54
                        |||   |||   :::               |||       |||
Db        1829 ------GCAGTCACAGCCAATTTGGATATTAGAAATGCTACAAAACTTAGAGTTTCTCCT 1882

Qy          55 LysGluProMetLeuProLysGluThrValLeuSerLeuLysArgTyrAsnLeuGlySer 74
               |||   |||   :::||||||      :::|||   |||         :::
Db        1883 AAGAATCCTCGTATCCCCAAATTGCATATGCTTGAATTACATTGTGAAAGCAAATGTGAC 1942

Qy          75 SerAlaMetLysArgAsnVal---------------------------------------- 81
               |||   :::|||   ::::::
Db        1943 TCACATTTGAAACACAGTTTGAAGTTGTCCTGGAGTAAAGATGGAGAAGCCTTTGAAATT 2002

Qy          82 -----------ProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnVal 97
                        |||   ::::::         :::   ||||||         |||
Db        2003 AATGGCACAGAAGATGGCAGGATAATTATTGATGGAGCTAATTTGACCATATCTAATGTA 2062

Qy          98 ThrLeu---------------LeuCysAsnSerSerAlaGluAlaLeu 108
               ||||||                    |||   |||:::      ||||||
Db        2063 ACTTTAGAGGACCAAGGTATTTACTGCTGTTCAGCTCATACTGCTCTA 2110

RESULT 12
US-09-252-991A-3473/c
; Sequence 3473, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 3473
;   LENGTH: 1302
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3473

Alignment Scores:
Pred. No.:              7.51          Length:        1302
Score:                  71.50         Matches:       29
Percent Similarity:     40.62%        Conservative:  10
Best Local Similarity:  30.21%        Mismatches:    24
Query Match:            13.07%        Indels:        33
DB:                     4             Gaps:          5

US-10-679-865A-2 (1-108) x US-09-252-991A-3473 (1-1302)

Qy           5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
               |||   :::      |||:::||||||      :::   |||:::          |||
Db         412 CTCGTCGTGCCTGAGCAGCAGGTTCAT------CAATACCTCGATGTGACTGCCGTC--- 362

Qy          25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
                                                        |||      :::||||||
Db         361 ----------------------------------------CCTGCGCCAGAGGCGCGT 344

Qy          45 LeuProLeuGlu---------AlaLeuLeuGlySerAsnLysGluProMetLeuProLys 61
               |||       |||       ||||||         :::|||
Db         343 CTCGTAGCCGAACAGGCCGCCGCGCTCGCGCAGCAGCGCGCGGATCCGTCG--------- 293

Qy          62 GluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
                                        :::||||||   :::|||            :::
Db         292 -------------------------CATCTCGGCCTCGCCGCC---GACGAACAGTTG 263

Qy          82 ProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnVal 97
               ||||||||||||||||||||||||   :::|||   ||||||   |||
Db         262 CCCGGCCATGTCCTCCAGCGACCAGAGCACCTGTTGCGGATCGTCGTA 215

RESULT 13
US-09-252-991A-3464
; Sequence 3464, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 3464
```

```
;   LENGTH: 1332
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3464

Alignment Scores:
Pred. No.:              7.78          Length:        1332
Score:                  71.50         Matches:       29
Percent Similarity:     40.62%        Conservative:  10
Best Local Similarity:  30.21%        Mismatches:    24
Query Match:            13.07%        Indels:        33
DB:                     4             Gaps:          5

US-10-679-865A-2 (1-108) x US-09-252-991A-3464 (1-1332)

Qy          5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
              |||   :::      |||:::||||||      :::   |||:::          |||
Db        946 CTCGTCGTGCCTGAGCAGCAGGTTCAT------CAATACCTCGATGTGACTGCCGTC--- 996

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
                                                    |||      :::||||||
Db        997 ----------------------------------------CCTGCGCCAGAGGCGCGT 1014

Qy         45 LeuProLeuGlu---------AlaLeuLeuGlySerAsnLysGluProMetLeuProLys 61
              |||       |||       ||||||                 :::|||
Db       1015 CTCGTAGCCGAACAGGCCGCCGCGCTCGCGCAGCAGCGCGCGGATCCGTCG--------- 1065

Qy         62 GluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
                                     :::||||||   :::|||            :::
Db       1066 --------------------------CATCTCGGCCTCGCCGCC---GACGAACAGTTG 1095

Qy         82 ProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnVal 97
              ||||||||||||||||||||||||   :::|||   ||||||   |||
Db       1096 CCCGGCCATGTCCTCCAGCGACCAGAGCACCTGTTGCGGATCGTCGTA 1143

RESULT 14
US-09-252-991A-3485/c
; Sequence 3485, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 3485
;   LENGTH: 1722
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3485

Alignment Scores:
Pred. No.:              11.5          Length:        1722
Score:                  71.50         Matches:       29
Percent Similarity:     40.62%        Conservative:  10
Best Local Similarity:  30.21%        Mismatches:    24
Query Match:            13.07%        Indels:        33
DB:                     4             Gaps:          5

US-10-679-865A-2 (1-108) x US-09-252-991A-3485 (1-1722)

Qy          5 LeuProIleGluGlyGlnGluValHisPhePheGluIleLeuGluSerGluCysValLeu 24
              |||   :::      |||:::||||||      :::   |||:::          |||
Db        582 CTCGTCGTGCCTGAGCAGCAGGTTCAT------CAATACCTCGATGTGACTGCCGTC--- 532

Qy         25 TyrThrGlyTrpIleGluArgAlaSerGlySerSerIleTyrProGluAlaLysAlaArg 44
                                                    |||      :::||||||
Db        531 ----------------------------------------CCTGCGCCAGAGGCGCGT 514

Qy         45 LeuProLeuGlu---------AlaLeuLeuGlySerAsnLysGluProMetLeuProLys 61
              |||       |||       ||||||                 :::|||
Db        513 CTCGTAGCCGAACAGGCCGCCGCGCTCGCGCAGCAGCGCGCGGATCCGTCG--------- 463

Qy         62 GluThrValLeuSerLeuLysArgTyrAsnLeuGlySerSerAlaMetLysArgAsnVal 81
                                     :::||||||   :::|||            :::
Db        462 --------------------------CATCTCGGCCTCGCCGCC---GACGAACAGTTG 433

Qy         82 ProGlyHisValLeuGlnArgProSerTyrLeuThrArgIleGlnVal 97
              ||||||||||||||||||||||||   :::|||   ||||||   |||
Db        432 CCCGGCCATGTCCTCCAGCGACCAGAGCACCTGTTGCGGATCGTCGTA 385

RESULT 15
US-09-902-540-2522
; Sequence 2522, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 2522
;   LENGTH: 3198
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-2522

Alignment Scores:
Pred. No.:              29.7          Length:        3198
Score:                  71.50         Matches:       24
Percent Similarity:     40.70%        Conservative:  11
Best Local Similarity:  27.91%        Mismatches:    25
Query Match:            13.07%        Indels:        26
DB:                     4             Gaps:          2

US-10-679-865A-2 (1-108) x US-09-902-540-2522 (1-3198)

Qy         39 ProGluAlaLysAlaArgLeuProLeuGluAlaLeuLeuGlySerAsnLysGluProMet 58
              |||       |||     |||   :::   ||||||           ||||||
Db       1083 CCACGACAAAGCGCTGGAGCACCTGCGCAAGGTCTTCTCCGTCGAGCCGGAGAACCTCGA 1142

Qy         59 LeuProLysGluThrValLeuSerLeuLysArg--------------------------- 69
                |||:::|||   |||   :::|||:::|||
Db       1143 CGCCCACGAGAAGGCGTATCAAATCTACGTCGCGTCCGGGAACATGGCCCAGGCATCGGA 1202

Qy         70 --------------------------TyrAsnLeuGlySerSerAlaMetLysArgAsn 80
                                         |||     |||     |||   :::
Db       1203 GCAGCTGCTGAATGTGCTCCGCCTGTGTACGCGCACGGCCGACGTGCAGCGCGCGCAGCC 1262

Qy         81 ValProGlyHis-ValLeuGlnArg--------------------ProSerTyrLeuTh 93
              |||||||||||| :::||||||                      |||::::::|||::
Db       1263 GTACCTGGCCACCATCCTCCAGGAGAATCCCGCCCATCCCGAGGTGCCCGCGTTTCTGTC 1322

Qy         93 rArgIleGlnValThr 98
              :   ::::::   |||
Db       1323 CGTGCTGCGCACGACG 1338


Search completed: September  1, 2005, 00:17:33
Job time : 3779 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2005, 23:15:23 ; Search time 3565 Seconds
(without alignments)
9203.759 Million cell updates/sec

Title: US-10-679-865A-1
Perfect score: 862
Sequence: 1 atgatgttgtgtttggagaa.....taattaaaacaaatcttgca 862

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	56.8	6.6	836	9	CC541575	CH240_422
	2	52.6	6.1	387	7	CO991529	CO991529 UMC-pd3ov
c	3	49.6	5.8	572	4	BG081672	BG081672 H3067G11-
c	4	49.4	5.7	555	8	AZ029451	AZ029451 RPCI-23-2
c	5	49.4	5.7	770	8	AZ560599	AZ560599 RPCI-23-1
c	6	48	5.6	446	2	AW988531	AW988531 ug08f06.y
c	7	48	5.6	448	2	BE137122	BE137122 ug59h07.y
c	8	48	5.6	517	5	BX516445	BX516445 BX516445
c	9	48	5.6	765	6	CD352950	CD352950 UI-M-GL0-
	10	44.8	5.2	697	1	AV331125	AV331125 AV331125
	11	44	5.1	949	9	CNS073SC	AL427906 clone BA0
	12	43.4	5.0	964	9	CNS02GJO	AL196413 Tetraodon
	13	43.2	5.0	650	7	CF369839	rg56g11.y
	14	43.2	5.0	2291	3	AK031977	AK031977 Mus muscu
c	15	42.6	4.9	834	8	BZ488508	BZ488508 BONJQ35TF
c	16	42.6	4.9	876	8	BH156057	BH156057 ENTRZ53TR
	17	42.6	4.9	879	8	BH163228	BH163228 ENTTH34TF
	18	42.6	4.9	905	8	BH149021	BH149021 ENTQB36TF
	19	42.6	4.9	1192	9	CG751653	CG751653 P046-1-F0
c	20	42	4.9	500	5	BP113810	BP113810 BP113810
	21	41.6	4.8	441	7	CF803914	CF803914 rg76h11.y
c	22	41.6	4.8	815	5	BQ177910	BQ177910 UI-M-ER0-
	23	41.6	4.8	922	7	CK405277	CK405277 AUF_IfSpn
c	24	41.4	4.8	386	1	AA131238	AA131238 zl31h06.s
c	25	41.4	4.8	511	1	AU039051	AU039051 AU039051
	26	41.4	4.8	863	9	CR192433	CR192433 Forward s
	27	41.2	4.8	368	6	CB481372	CB481372 jns76_G05
c	28	41.2	4.8	574	5	BU276213	BU276213 Cr_Emb_10
c	29	41.2	4.8	959	9	CNS017OR	AL108309 Drosophil
c	30	41	4.8	655	7	CF111724	CF111724 Shultzomi
c	31	41	4.8	738	8	AQ582080	AQ582080 RPCI-11-4
c	32	41	4.8	744	9	CL652679	CL652679 PRI0115c_
	33	40.8	4.7	236	2	BB048122	BB048122 BB048122
c	34	40.8	4.7	673	2	BF298919	BF298919 065PbA03
	35	40.8	4.7	705	8	AZ381766	AZ381766 1M0138A04
	36	40.8	4.7	756	8	BZ499202	BZ499202 BONKB08TF
	37	40.6	4.7	619	9	CG187363	CG187363 PUIFD76TB
c	38	40.6	4.7	669	5	BP516340	BP516340 BP516340
	39	40.6	4.7	744	9	CG186597	CG186597 PUIFD75TB
	40	40.6	4.7	833	9	CG066711	CG066711 PUFSR11TB
	41	40.6	4.7	980	1	AL514519	AL514519 AL514519
c	42	40.6	4.7	1101	9	CNS00GEB	AL072080 Drosophil
c	43	40.4	4.7	510	9	CNS00BK6	AL057280 Drosophil
c	44	40.4	4.7	683	9	CE273727	CE273727 tigr-gss-
c	45	40.4	4.7	1101	9	CNS0027I	AL062252 Drosophil

ALIGNMENTS

RESULT 1
CC541575
LOCUS CC541575 836 bp DNA linear GSS 17-JUN-2003
DEFINITION CH240_422N19.T7 CHORI-240 Bos taurus genomic clone CH240_422N19, genomic survey sequence.
ACCESSION CC541575
VERSION CC541575.1 GI:31859859
KEYWORDS GSS.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 836)
AUTHORS Holt,R., Stott,J., Yang,G., Barber,S., Smailus,D., Prabhu,A.-L., Tsai,M., Cloutier,A., Lee,D., Girn,N., Olson,T., Mayo,M., Butterfield,Y., Kirkpatrick,R., Liu,J., Guin,R., Chan,A., Chiu,R., Mathewson,C., Wye,N., Masson,A., Brown-John,M., Jones,S., Schein,J., Marra,M., de Jong,P., Keele,J.W. and Kappes,S.M.
TITLE Bovine BAC End Sequences from Library CHORI-240, PLATES 399 to 478
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: CH240_422N19.TARBAC13P2
Contact: Rob Holt
Sequencing
The British Columbia Cancer Agency Genome Science Centre
600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6
Tel: 604-877-6085
Fax: 604-877-6276
Email: rholt@bcgsc.ca
Clones are derived from the bovine BAC library CHORI-240 (http://www.chori.org/bacpac/bovine240.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/ordering_information.htm). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the British Columbia Genome Sciences Centre, Canada.
Plate: 422 row: N column: 19
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .836
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_422N19"

/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI; Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 6.6%; Score 56.8; DB 9; Length 836;
Best Local Similarity 73.5%; Pred. No. 0.00078;
Matches 100; Conservative 0; Mismatches 32; Indels 4; Gaps 2;

```
Qy          2 TGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAAG 61
              ||||  |     ||||   ||||||| ||| ||| | |||| ||||||||||||||||||
Db        100 TGATAGTCCACATGGAAGTAAGAGAGGAGAAAATAGCATAAATACATGACCCCTTGGAAG 159

Qy         62 CTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAAT 121
              ||||||||   || |||  |   ||||||| || |||||||||||| ||| |  | |||
Db        160 CTGTTTTC---TGTGGAAATATTTGAGTAGGCAGTTGAAATCAAGGTAAA-TATAGGAAG 215

Qy        122 AGCAGGAGCTGGGAAG 137
              |  ||||| ||||| |
Db        216 AATAGGAGTTGGGAGG 231
```

RESULT 2
CO991529
LOCUS CO991529 387 bp mRNA linear EST 09-SEP-2004
DEFINITION UMC-pd3ov4-002-a08 Oviduct gilt D3 of estrous cycle pd3ov Sus scrofa cDNA 3', mRNA sequence.
ACCESSION CO991529
VERSION CO991529.1 GI:51350803
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 387)
AUTHORS Kim,J., Whitworth,K.M., Springer,G.K., Bivens,N.J., Ries,J.E., Woods,R.J., Spollen,W.G., Forrester,L.J., Mathialagan,N., Prather,R.S. and Green,J.A.
TITLE Large-scale Generation and Analysis of Expressed Sequence Tags from Porcine endometrium and oviduct
JOURNAL Unpublished (2004)
COMMENT Contact: DNA Core Facility (Swine Project)
Animal Science - RS Prather
University of Missouri-Columbia
M616 Medical Sciences Bldg., Columbia, MO 65212, USA
Tel: (573)882-0428
Fax: (573)884-5552
Email: porcine@rnet.missouri.edu
POLYA=No.
FEATURES Location/Qualifiers
source 1. .387
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/dev_stage="Oviduct from a gilt on day 3 of the estrous cycle"
/clone_lib="pd3ov"
/note="Vector: pSport1; Funding: A grant from the Monsanto Company to the University of Missouri. Genetic Source: Endometrium and oviduct tissues from various stages of the estrous cycle were collected from crossbred pigs (Sus scrofa domestica), frozen in liquid nitrogen immediately after collection, and stored at -80 degrees Celsius until RNA extraction. The specific tissues collected were Day 0 and Day 3 whole oviducts and Days 3, 6, 10 and 12-14 endometrium. More information regarding the methods can be found at:
http://genome.rnet.missouri.edu/Swine/Methods.html.
Library Construction (Standard Protocol): All procedures discussed in this section have been described in detail elsewhere (Soares et al., 1994; Bonaldo et al., 1996; Jiang et al., 2001). Total cellular RNA from each sample was isolated by using STAT-60 reagent (Tel-Test, Friendswood, TX) and poly(A)+ RNA was obtained by two rounds of purification with the Oligotex mRNA isolation kit (Qiagen) according to the manufacturer's instructions. The oviduct libraries and the Day 3, 6 and 10 endometrium libraries were constructed essentially as described by the manufacturer's instructions provided with the SuperScript Plasmid System (Invitrogen, cat. no. 18248-013). Briefly, 1mg of poly(A)+ RNA will be annealed at 37 degrees Celsius with 10mg of NotI-tag-dT18 oligonucleotide (GCTGCTCGCGGCCGC-tag-T18) and reverse transcribed at 37 degrees Celsius with SuperScript II (Invitrogen) reverse transcriptase (Jiang et al., 2001). The 'tag' represents a tissue/stage-specific ten-base sequence identifier (http://genome.uiowa.edu/pubsoft/software.html) present in the oligonucleotide used to prime first-strand synthesis. Second strand synthesis was performed with T4 DNA polymerase in the presence of DNA ligase and RNase H. After second strand synthesis, the double-stranded cDNAs was ligated to SalI adapters (Invitrogen-Life Technologies) and digested with NotI. The cDNAs will be size selected by passage through cDNA size fractionation columns (Invitrogen-Life technologies). The cDNAs derived from each developmental stage of a particular tissue were mixed on an equimolar basis and ligated directionally into the NotI and SalI sites of the pSPORT1 vector (Invitrogen). After ligation of the inserts, the plasmids will be electroporated into DH10B bacteria. The day 12-14 endometrium library was synthesized by Dr. Bento Soares' laboratory (University of Iowa) and was cloned into the T3T7pac vector as described elsewhere (Bonaldo et al., 1996). Preliminary Library Characterization: Randomly chosen clones from each library were analyzed by restriction digestion to determine average insert size (96 clones) and by sequencing (~4 96-well plates) to confirm library quality [e.g. the presence of short polyA+ tails, genomic DNA contamination (must be <1%), ribosomal RNA clones (must be <1%), etc.] and to provide a sequence database representing the predominant clones in each library. The clones were sequenced at the University of Missouri-Columbia DNA Core Facility. Bioinformatics work was performed by GK Springer's bioinformatics group (WG Spollen, JE Ries, A Guillen, AA Khambati, RV Patel, CM Topinka, SB Bhuiyan) in Computer Science and Health Management and Informatics Departments at the University of Missouri-Columbia. Clone Requests: Requests for clones should be made to the Director of the University of Missouri DNA Core facility at: porcine@rnet.missouri.edu. Citations: 1. Bonaldo MF, Lennon G, Soares MB. Normalization and Subtraction: Two approaches to facilitate gene discovery. Genome Res, 1996; 6:791-806. 2. Jiang H, Bivens NJ, Ries JE, Whitworth KM, Green JA, Forrester LJ, Springer GK, Didion BA, Mathialagan N, Prather RS, Lucy MC (2001) Constructing cDNA libraries with fewer clones that contain long poly(dA) tails. Biotechniques 31:38-42. 3. Soares MB, MF Banaldo, P Jelene, L Su, L Lawton, A Efstrantiadis. 1994. Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci, 91:9228-9232.
TAG_TISSUE=Oviduct gilt D3 of estrous cycle
TAG_SEQ=Not found"

ORIGIN

Query Match 6.1%; Score 52.6; DB 7; Length 387;
Best Local Similarity 69.9%; Pred. No. 0.0084;
Matches 100; Conservative 0; Mismatches 39; Indels 4; Gaps 2;

```
Qy          3 GATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAAGC 62
              |||| | |  |   || |         | | ||| | |||| || | ||||| ||||||||
Db        242 GATGATATTCTACAAGGAGGTGAGAGAAAAAATGGCATAAATATAGGACCCTTTGGAAGC 301
```

```
Qy         63 TGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAATA 122
              ||||||||||||| ||||||   ||||||| |||||||||| |  ||||| |  | |||||
Db        302 TGTTTTCAGTTGGGGATTT---TGAGTAGGCACTTGAAATTACAGGAAA-TATAGGAATA 357

Qy        123 GCAGGAGCTGGGAAGAAAGGAAG 145
              | ||||| |||||| |   || |
Db        358 GTAGGAGTTGGGAAAAGGAGAGG 380
```

RESULT 3
BG081672/c
LOCUS BG081672 572 bp mRNA linear EST 17-DEC-2003
DEFINITION H3067G11-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone H3067G11 5', mRNA sequence.
ACCESSION BG081672
VERSION BG081672.2 GI:40016442
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 572)
AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H. III, Becker,K.G. and Ko,M.S.H.
TITLE Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
MEDLINE 20381348
PUBMED 10922068
COMMENT On Jan 26, 2001 this sequence version replaced gi:12564240.
Other_ESTs: H3067G11-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3067 row: G column: 11
Seq primer: -21M13 Reverse
High quality sequence stop: 572
POLYA=No.
FEATURES Location/Qualifiers
source 1. .572
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:H3067G11-5"
/db_xref="taxon:10090"
/clone="H3067G11"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse develolpment, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
ORIGIN

Query Match 5.8%; Score 49.6; DB 4; Length 572;
Best Local Similarity 55.1%; Pred. No. 0.058;
Matches 97; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```
Qy        684 CACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCC 743
              | |  || ||| | ||||| |||||  ||  ||||  ||||  |  |||  | | |
Db        244 CTCTGCAGAGATGTAAACTGGATTTTTATTTGACCATATTCACACTAGTAATATTTGTAA 185

Qy        744 ACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTT 803
              |||| ||   | ||| | |  || ||   ||| || ||||   | ||||| | || |
Db        184 ACTTATATTTAAAACGTATTAAACTTATTTTCTAAATGTGAGTATAAACATTGAATTCGG 125

Qy        804 TAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTT 859
              || |  ||| | |     | |    ||| |  |   ||| |   ||  | |||||
Db        124 TATACTGGAAGAGCCAATAAATACTCAGTCACAGAACTTTTACCTAATTATATCTT 69
```

RESULT 4
AZ029451/c
LOCUS AZ029451 555 bp DNA linear GSS 25-FEB-2000
DEFINITION RPCI-23-244P10.TV RPCI-23 Mus musculus genomic clone RPCI-23-244P10, genomic survey sequence.
ACCESSION AZ029451
VERSION AZ029451.1 GI:7104835
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 555)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-244P10.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 244 row: P column: 10
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .555
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-244P10"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "

ORIGIN

Query Match 5.7%; Score 49.4; DB 8; Length 555;
Best Local Similarity 54.0%; Pred. No. 0.065;
Matches 101; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```
Qy        669 CCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAA 728
              || | | | ||   | | |||| ||||  | |  || || | || |||   |||    |
Db        307 CCTTAAGTTAGAAAGTAAAACAGAGAAATACATATGTTTCCTATTAGAAAAAATGGAGAG 248

Qy        729 CAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAG 788
              |   ||||||    | ||| |       |||   ||   || |   |||  ||  | |
Db        247 AATCATTTTCTGATAAATTTCAAGATGTTCTTAAACAGCATTTGGATAAAAAGTAAAAAA 188

Qy        789 AAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTA 848
                |||| |||  |   |||| ||| ||||| | |||  ||||| ||   | |   |   |
Db        187 TTACAAAAAACGTGGAAAAATGGAGAAACTTGATAGAAAAAAGAAACCTGGACAAATGGA 128

Qy        849 AAACAAA 855
              |||||  |
Db        127 AAACAGA 121
```

RESULT 5
AZ560599/c
LOCUS AZ560599 770 bp DNA linear GSS 20-NOV-2000
DEFINITION RPCI-23-173A21.TV RPCI-23 Mus musculus genomic clone RPCI-23-173A21, genomic survey sequence.
ACCESSION AZ560599
VERSION AZ560599.1 GI:11240419
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 770)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-173A21.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 173 row: A column: 21
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .770
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-173A21"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "

ORIGIN

Query Match 5.7%; Score 49.4; DB 8; Length 770;
Best Local Similarity 54.0%; Pred. No. 0.071;
Matches 101; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```
Qy        669 CCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAA 728
              || | | | ||   | | |||| ||||  | |  || || | || |||   |||    |
Db        622 CCTTAAGTTAGAAAGTAAAACAGAGAAATACATATGTTTCCTATTAGAAAAAATGGAGAG 563

Qy        729 CAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAG 788
              |   ||||||    | ||| |       |||   ||   || |   |||  ||  | |
Db        562 AATCATTTTCTGATAAATTTCAAGATGTTCTTAAACAGCATTTGGATAAAAAGTAAAAAA 503

Qy        789 AAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTA 848
                |||| |||  |   |||| ||| ||||| | |||  ||||| ||   | |   |   |
Db        502 TTACAAAAAACGTGGAAAAATGGAGAAACTTGATAGAAAAAAGAAACCTGGACAAATGGA 443

Qy        849 AAACAAA 855
              |||||  |
Db        442 AAACAGA 436
```

RESULT 6
AW988531/c
LOCUS AW988531 446 bp mRNA linear EST 02-JUN-2000
DEFINITION ug08f06.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone IMAGE:1531043 5', mRNA sequence.
ACCESSION AW988531
VERSION AW988531.1 GI:8183702
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 446)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:947143
Seq primer: -40RP from Gibco
High quality sequence stop: 438.
FEATURES Location/Qualifiers
source 1. .446
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:1531043"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/clone_lib="Soares_mammary_gland_NMLMG"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 5.6%; Score 48; DB 2; Length 446;
Best Local Similarity 54.5%; Pred. No. 0.14;
Matches 96; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy        684 CACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCC 743
              | |  || ||| | ||||| |||||  ||  ||||  ||||  |  |||  | | |
Db        264 CTCTGCAGAGATGTAAACTGGATTTTTATTTGACCATATTCACACTAGTAATATTTGTAA 205

Qy        744 ACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTT 803
              ||||  ||   | ||| | |  || ||   ||| || ||||   |  |||| | || |
Db        204 ACTTATATTTAAAACGTATTAAACTTATTTTCTAAATGTGAGTATCAACATTGAATTCGG 145

Qy        804 TAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTT 859
              ||  |  |||  | |     | |    ||| |  |   ||| |   ||  | |||||
Db        144 TATACTGGAAGAGCCAATAAATACTCAGTCACAGAACTTTTACCTAATTATATCTT 89
```

RESULT 7
BE137122/c

LOCUS	BE137122 448 bp mRNA linear EST 21-JUN-2000
DEFINITION	ug59h07.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone IMAGE:1546717 5', mRNA sequence.
ACCESSION	BE137122
VERSION	BE137122.1 GI:8599622
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 448)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:952065 Seq primer: -40RP from Gibco High quality sequence stop: 416.

FEATURES Location/Qualifiers

source 1. .448
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:1546717"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/clone_lib="Soares_mammary_gland_NMLMG"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 5.6%; Score 48; DB 2; Length 448;
Best Local Similarity 54.5%; Pred. No. 0.14;
Matches 96; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy        684 CACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCC 743
              | |  || ||| | ||||| |||||  ||  ||||  ||||  |  |||  | | |
Db        264 CTCTGCAGAGATGTAAACTGGATTTTTATTTGACCATATTCACACTAGTAATATTTGTAA 205

Qy        744 ACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTT 803
              ||||  ||   | ||| | |  || ||   ||| || ||||   |  |||| | || |
Db        204 ACTTATATTTAAAACGTATTAAACTTATTTTCTAAATGTGAGTATCAACATTGAATTCGG 145

Qy        804 TAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTT 859
              ||  |  |||  | |     | |    ||| |  |   ||| |   ||  | |||||
Db        144 TATACTGGAAGAGCCAATAAATACTCAGTCACAGAACTTTTACCTAATTATATCTT 89
```

RESULT 8
BX516445/c

LOCUS	BX516445 517 bp mRNA linear EST 27-JUN-2003
DEFINITION	BX516445 Soares_mammary_gland_NMLMG Mus musculus cDNA clone IMAGp998D143918 ; IMAGE:1546717, mRNA sequence.
ACCESSION	BX516445
VERSION	BX516445.1 GI:32298939
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 517)
AUTHORS	Heil,O., Ebert,L., Neubert,P., Peters,M., Radelof,U., Schneider,D. and Korn,B.
TITLE	Mouse UnigeneSet - RZPD2
JOURNAL	Unpublished (2003)
COMMENT	Contact: Ina Rolfs RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany RZPD; IMAGp998D143918. RZPDLIB; I.M.A.G.E. cDNA Clone Collection; Mouse UnigeneSet - RZPD2 (RZPDLIB No.981) http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi/response?libNo=981 Contact: Ina Rolfs RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Heubnerweg 6, D-14059 Berlin, Germany Tel: +49 30 32639 101 Fax: +49 30 32639 111 www.rzpd.de This clone is available royalty-free from RZPD; contact RZPD (clone@rzpd.de) for further information. Seq primer: T7, Primer sequence: TAATACGACTCACTATAGGG.

FEATURES Location/Qualifiers

source 1. .517
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGp998D143918 ; IMAGE:1546717"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/clone_lib="Soares_mammary_gland_NMLMG"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 5.6%; Score 48; DB 5; Length 517;
Best Local Similarity 54.5%; Pred. No. 0.15;
Matches 96; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy        684 CACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCC 743
              | |  || ||| | ||||| |||||  ||  ||||  ||||  |  |||  | | |
Db        265 CTCTGCAGAGATGTAAACTGGATTTTTATTTGACCATATTCACACTAGTAATATTTGTAA 206

Qy        744 ACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTT 803
              ||||  ||   | ||| | |  || ||   ||| || ||||   |  |||| | || |
Db        205 ACTTATATTTAAAACGTATTAAACTTATTTTCTAAATGTGAGTATCAACATTGAATTCGG 146

Qy        804 TAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTT 859
              ||  |  |||  | |     | |    ||| |  |   ||| |   ||  | |||||
Db        145 TATACTGGAAGAGCCAATAAATACTCAGTCACAGAACTTTTACCTAATTATATCTT 90
```

RESULT 9
CD352950/c

```
LOCUS       CD352950                765 bp    mRNA    linear   EST 15-JUL-2003
DEFINITION  UI-M-GL0-cfz-f-09-0-UI.r1 NIH_BMAP_GL0 Mus musculus cDNA clone
            IMAGE:30359648 5', mRNA sequence.
ACCESSION   CD352950
VERSION     CD352950.1  GI:31145451
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 765)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
             The following repetitive elements were found in this cDNA
            sequence: 687-713, >AT_rich#Low_complexity
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .765
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30359648"
                     /tissue_type="Whole brain"
                     /dev_stage="1, 5 and 15 days newborn"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_GL0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel.First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site.Double strand cDNA was size
                     selected according to mRNA size fraction ,ligated with
                     EcoR I adaptor , digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is CGAACTGAAT. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin , Ph.D. ,
                     program coordinator."
ORIGIN
```

```
  Query Match             5.6%;  Score 48;  DB 6;  Length 765;
  Best Local Similarity   54.5%;  Pred. No. 0.17;
  Matches   96;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        684 CACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCC 743
              | |  || ||| | ||||| |||||  ||  ||||  ||||  |  |||  | | |
Db        269 CTCTGCAGAGATGTAAACTGGATTTTTATTTGACCATATTCACACTAGTAATATTTGTAA 210

Qy        744 ACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTT 803
              |||| ||   | ||| | |  || ||   ||| || ||||   |  |||| | || |
Db        209 ACTTATATTTAAAACGTATTAAACTTATTTTCTAAATGTGAGTATCAACATTGAATTCGG 150

Qy        804 TAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTT 859
              || |  |||  | |     | |   ||| |  |  |||  |   ||  | |||||
Db        149 TATACTGGAAGAGCCAATAAATACTCAGTCACAGAACTTTTACCTAATTATATCTT 94
```

```
RESULT 10
AV331125
LOCUS       AV331125                697 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  AV331125 RIKEN full-length enriched, adult male medulla oblongata
            Mus musculus cDNA clone 6330520K10 3', mRNA sequence.
ACCESSION   AV331125
VERSION     AV331125.2  GI:16396168
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 11, 1999 this sequence version replaced gi:6371177.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6330520K10"
                     /sex="male"
                     /tissue_type="medulla oblongata"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male medulla
                     oblongata"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
```

prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

ORIGIN

Query Match 5.2%; Score 44.8; DB 1; Length 697;
Best Local Similarity 55.0%; Pred. No. 1.1;
Matches 88; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```
Qy        702 TTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTT 761
              ||  |||  | ||||  || | |||  |  ||    |  || ||||||| |||  ||
Db        528 TTTCTTTTCAAAAGAATGAAGTTTTAGAATGATTAAAGAGTCAGTTTTACTCATGCATTA 587

Qy        762 TCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGC 821
              ||||||||  ||  || |    | |  ||||||  |||  |     ||| ||     ||
Db        588 CCCAAATTTTCTTGCAAGTTGCAAAAAGAACAATGTAATGATATGTGAGGTTACTTATGC 647

Qy        822 TAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTTGC 861
              || | ||| |    |     | | ||||||||||  || |
Db        648 TAATAAAATGATCTTTTAGATGAGTTAAAACAAAAGTTCC 687
```

RESULT 11
CNS073SC
LOCUS CNS073SC 949 bp DNA linear GSS 07-JUL-2001
DEFINITION clone BA0AB022E02 of library BA0AB from strain CLIB 210 of Kluyveromyces lactis, genomic survey sequence.
ACCESSION AL427906
VERSION AL427906.1 GI:12211100
KEYWORDS GSS.
SOURCE Kluyveromyces lactis
ORGANISM Kluyveromyces lactis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE 1 (bases 1 to 949)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 949)
AUTHORS Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F., Duchateau-Nguyen,G., Lemaire,M., Marmeisse,R., Montrocher,R., Robert,C., Termier,M., Wincker,P. and Wesolowski-Louvel,M.
TITLE Genomic exploration of the hemiascomycetous yeasts: 11. Kluyveromyces lactis
JOURNAL FEBS Lett. 487 (1), 66-70 (2000)
MEDLINE 20584721
PUBMED 11152886
REFERENCE 3 (bases 1 to 949)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES Location/Qualifiers
source 1. .949
/organism="Kluyveromyces lactis"
/mol_type="genomic DNA"
/strain="CLIB 210"
/variety="lactis"
/db_xref="taxon:28985"
/clone="BA0AB022E02"
/clone_lib="BA0AB"

ORIGIN

Query Match 5.1%; Score 44; DB 9; Length 949;
Best Local Similarity 41.8%; Pred. No. 2;
Matches 69; Conservative 23; Mismatches 73; Indels 0; Gaps 0;

```
Qy        694 AAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCC 753
              ||  ||| ||   ||  |||    | |   || :|   |  ||| :       ||    :
Db        265 AANAAAAATTTTATTTNATATATACAATAACTAWAGKTTTATTTAKTTTTNNCTTNKNNH 324

Qy        754 ATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGAT 813
               ::  :||: || :||   |    ||| | |: : ::|:|: ::|:|   ||||||  |
Db        325 GYMCAWTTYTCANWTTCNTTNATNATTATTCHAHMMHAMAMMMAWTAAATAAAAAAAAA 384

Qy        814 AAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCT 858
              ||::|  :|:  |: | |  ||    |||||:|||| ||||
Db        385 AAMMTAAMTWHATWTATTTTTATATTTTTTAAWTAAATTAAATTT 429
```

RESULT 12
CNS02GJO
LOCUS CNS02GJO 964 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 137D24 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL196413
VERSION AL196413.1 GI:7834563
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143

```
REFERENCE   3  (bases 1 to 964)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .964
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="137D24"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG137DB12SP1~end :
                     PUC-Ori"
ORIGIN

  Query Match             5.0%;  Score 43.4;  DB 9;  Length 964;
  Best Local Similarity   44.4%;  Pred. No. 2.9;
  Matches   75;  Conservative   21;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        687 AACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACT 746
              :| : | ||: |||     |    |:| :     ||:  : |: | | |:|:  ||:|| |
Db        757 RARRAAAAARAAAAAAAAAAAAAAWTTWWTTTWAAWAWWAAWAAATTTKTKGAACHCAAT 816

Qy        747 TTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAA 806
              |||    | ||   :|  |: |||  ||     || | | | ||| || ||||   |::|
Db        817 TTTTTTTAAAAAAAWTAAAWTTTTTTTTTTTTGTGAGAAAAAAAAAAAAAAAAAAAWWA 876

Qy        807 AAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAA 855
              |||  |::: :    |: || ||: |   |    |||| |||||| |||
Db        877 AAAWAAWWWTMAAATTWWTTTAAWTTTTTTTTTTTTTATTTAAAAAAAA 925


RESULT 13
CF369839
LOCUS       CF369839                650 bp    mRNA    linear   EST 27-AUG-2003
DEFINITION  rg56g11.y1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla
            cDNA 5' similar to TR:Q9VQT7 Q9VQT7 CG16713 PROTEIN. ;, mRNA
            sequence.
ACCESSION   CF369839
VERSION     CF369839.1  GI:34317085
KEYWORDS    EST.
SOURCE      Meloidogyne hapla
  ORGANISM  Meloidogyne hapla
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 650)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified
            using Dynabeads (Dynal) and mRNA eluted for first strand synthesis.
            First strand cDNA was created using MMLV RT (Powerscript, Clontech)
            and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added
            using chimeric DNA-RNA oligo. 12 PCR cycles were done using first
            strand and primers specific to SMART oligo and 3' end. Double
            stranded cDNA was digested using XhoI/NotI, fractioned on
            Chroma-spin 400 columns (Clontech) and ligated to digested
            pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as
            host cells. Library materials provided by Dr. David Bird of North
            Carolina State University. Library construction by Jeff Rousch. See
            www.nematode.net for additional project information.
            Seq primer: Sp6.
FEATURES             Location/Qualifiers
     source          1. .650
                     /organism="Meloidogyne hapla"
                     /mol_type="mRNA"
                     /db_xref="taxon:6305"
                     /tissue_type="whole organism"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Meloidogyne hapla female SMART pGEM"
                     /note="Vector: plasmid (ampicillin resistant); Site_1:
                     XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA
                     was concentrated and purified using Dynabeads (Dynal) and
                     mRNA eluted for first strand synthesis. First strand cDNA
                     was created using MMLV RT (Powerscript, Clontech) and
                     primed with oligo(dT) with XhoI site and 5'SMART 'anchor'
                     added using chimeric DNA-RNA oligo. 12 PCR cycles were
                     done using first strand and primers specific to SMART
                     oligo and 3' end. Double stranded cDNA was digested using
                     XhoI/NotI, fractioned on Chroma-spin 400 columns
                     (Clontech) and ligated to digested pGEM-11zf(+) plasmid.
                     Chemically competent DH10B cells were used as host cells.
                     Library materials provided by Dr. David Bird of North
                     Carolina State University. Library construction by Jeff
                     Rousch. See www.nematode.net for additional project
                     information."
ORIGIN

  Query Match             5.0%;  Score 43.2;  DB 7;  Length 650;
  Best Local Similarity   50.0%;  Pred. No. 3;
  Matches  108;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        640 ATGCAGCAAGATATTAGGACAGATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAA 699
              ||  ||| | || | || |    ||        || ||| | | || | ||   |  | |||
Db        237 ATAAAGCTATATGTGAGCAAGAATGCAAGAAATAATTGATGATGAAGAATGTAAGAAGAA 296

Qy        700 ACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACT 759
              | || |||  ||  | |    ||| |||| |  |  || ||  | |||       |||
Db        297 AGTTAATTAAAAAGAAATAATATTATTAATATACAATTATATACGCTTAATTGTGTAATA 356

Qy        760 TTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTG 819
              | |  ||| |         ||  |  | | ||| || ||||     |||||  | |||
Db        357 TGTTGAAAATAAACATTTATAATTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 416

Qy        820 GCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAA 855
                |   ||||  |||    |   ||  |||| |||
Db        417 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 452


RESULT 14
AK031977
LOCUS       AK031977                2291 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus adult male medulla oblongata cDNA, RIKEN full-length
            enriched library, clone:6330520K10 product:phosphodiesterase 4B,
            cAMP specific, full insert sequence.
ACCESSION   AK031977
VERSION     AK031977.1  GI:26082622
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
```

MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2291)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES Location/Qualifiers
source 1. .2291
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:6330520K10"
/db_xref="taxon:10090"
/clone="6330520K10"
/sex="male"
/tissue_type="medulla oblongata"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
misc_feature 1. .2291
/note="phosphodiesterase 4B, cAMP specific (MGD|MGI:99557, GB|NM_019840, evidence: BLASTN, 99%, match=1104)"
ORIGIN

Query Match 5.0%; Score 43.2; DB 3; Length 2291;
Best Local Similarity 54.4%; Pred. No. 4.1;
Matches 87; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy        702 TTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTT 761
              || ||| | ||||  || | ||| |  ||    |  || ||||||| |||  ||
Db       2122 TTTCTTTTCAAAAGAATGAAGTTTTAGAATGATTAAAGAGTCAGTTTTACTCATGCATTA 2181

Qy        762 TCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGC 821
              ||||||| ||  || | | | |||||| ||| |    ||| ||    ||
Db       2182 CACAAATTTTCTTGCAAGTTGCAAAAAGAACAATGTAATGATATGTGAGGTTACTTATGC 2241

Qy        822 TAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTTGC 861
              || | ||| |   |    | | |||||||||| || |
Db       2242 TAATAAAATGATCTTTTAGATGAGTTAAAACAAAAGTTCC 2281
```

RESULT 15
BZ488508/c
LOCUS BZ488508 834 bp DNA linear GSS 16-DEC-2002
DEFINITION BONJQ35TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BONJQ35, genomic survey sequence.
ACCESSION BZ488508
VERSION BZ488508.1 GI:26995058
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 834)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BONJQ35TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .834
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BONJQ35"
/clone_lib="BO_1.6_2_KB_tot"
/note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Query Match 4.9%; Score 42.6; DB 8; Length 834;
Best Local Similarity 61.1%; Pred. No. 4.6;
Matches 69; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

```
Qy        747 TTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAA 806
              ||||  || | |||| |||||||  |  |||||||| |   ||| |  ||| || |||
Db        827 TTTAATAATCAGTTTTACAAATTTAATGAAAATTGTGAACTAAAATATAAAATTTCTTAG 768

Qy        807 AAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTT 859
                || |  ||  | |  | |||||  |  |||   ||||  ||| | | |. ||
```

Db 767 TAATGCAAATTTTGAAATTTAAATATTTATGCATTTTATATAATATATAGTTT 715

Search completed: September 1, 2005, 02:28:33
Job time : 3572 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2005, 00:10:13 ; Search time 201 Seconds
(without alignments)
7017.266 Million cell updates/sec

Title: US-10-679-865A-1
Perfect score: 862
Sequence: 1 atgatgttgtgtttggagaa.....taattaaaacaaatcttgca 862

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	41.8	4.8	1141	4	US-09-806-708B-22	Sequence 22, Appl
c	2	41.6	4.8	477	4	US-09-248-796A-742	Sequence 742, App
	3	39.4	4.6	105050	4	US-09-949-016-15953	Sequence 15953, A
c	4	38.4	4.5	832	4	US-09-621-976-2813	Sequence 2813, Ap
	5	38.2	4.4	99916	4	US-09-816-095-3	Sequence 3, Appli
c	6	37.4	4.3	7218	1	US-08-232-463-14	Sequence 14, Appl
	7	37.2	4.3	640681	4	US-09-790-988-1	Sequence 1, Appli
c	8	37	4.3	37715	4	US-09-949-016-13846	Sequence 13846, A
	9	36.8	4.3	1141	4	US-09-806-708B-22	Sequence 22, Appl
	10	36.4	4.2	157644	4	US-09-949-016-16179	Sequence 16179, A
	11	36.4	4.2	157644	4	US-09-949-016-16180	Sequence 16180, A
	12	36.2	4.2	601	4	US-09-949-001-379	Sequence 379, App
	13	36.2	4.2	312957	4	US-09-949-001-31	Sequence 31, Appl
	14	36.2	4.2	312972	4	US-09-949-001-34	Sequence 34, Appl
c	15	36	4.2	194933	4	US-09-949-016-14172	Sequence 14172, A
c	16	36	4.2	636591	4	US-09-949-016-11808	Sequence 11808, A
c	17	36	4.2	636591	4	US-09-949-016-13388	Sequence 13388, A
c	18	35.8	4.2	474	4	US-09-621-976-18033	Sequence 18033, A
c	19	35.8	4.2	4140	3	US-08-894-731-2	Sequence 2, Appli
c	20	35.6	4.1	32594	4	US-09-949-016-12219	Sequence 12219, A
c	21	35.6	4.1	32658	4	US-09-949-016-16242	Sequence 16242, A
	22	35.4	4.1	528	3	US-09-308-003-43	Sequence 43, Appl
	23	35.4	4.1	990	3	US-09-308-003-4	Sequence 4, Appli
	24	35.4	4.1	12173	4	US-08-956-171E-310	Sequence 310, App
	25	35.4	4.1	12173	4	US-08-781-986A-310	Sequence 310, App
	26	35.2	4.1	207	4	US-09-248-796A-10490	Sequence 10490, A
c	27	35.2	4.1	18303	4	US-09-949-016-13628	Sequence 13628, A
c	28	35	4.1	601	4	US-09-949-016-189977	Sequence 189977,
	29	35	4.1	6881	4	US-09-949-016-17181	Sequence 17181, A
c	30	35	4.1	640681	4	US-09-790-988-1	Sequence 1, Appli
c	31	34.6	4.0	150032	4	US-09-949-016-14321	Sequence 14321, A
c	32	34.4	4.0	399	4	US-09-621-976-8976	Sequence 8976, Ap
c	33	34.4	4.0	601	4	US-09-949-016-144587	Sequence 144587,
c	34	34.4	4.0	99830	4	US-09-949-016-16859	Sequence 16859, A
c	35	34.4	4.0	192506	4	US-09-949-016-15830	Sequence 15830, A
	36	34.2	4.0	807	4	US-09-248-796A-7603	Sequence 7603, Ap
	37	34.2	4.0	94593	4	US-09-949-016-16324	Sequence 16324, A
	38	34.2	4.0	95890	4	US-09-949-016-16412	Sequence 16412, A
c	39	34.2	4.0	113701	4	US-09-949-016-13214	Sequence 13214, A
c	40	34	3.9	60417	4	US-09-949-016-13312	Sequence 13312, A
c	41	34	3.9	129899	4	US-09-949-016-14684	Sequence 14684, A
c	42	34	3.9	148156	4	US-09-949-016-11776	Sequence 11776, A
c	43	34	3.9	205163	4	US-09-949-016-17009	Sequence 17009, A
c	44	33.8	3.9	1442	4	US-09-270-767-6007	Sequence 6007, Ap
c	45	33.8	3.9	1442	4	US-09-270-767-21289	Sequence 21289, A

ALIGNMENTS

RESULT 1
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 4.8%; Score 41.8; DB 4; Length 1141;
Best Local Similarity 10.0%; Pred. No. 0.082;
Matches 66; Conservative 261; Mismatches 333; Indels 3; Gaps 1;

```
Qy      197 CAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGAATAAAGCGAATGGAAG 256
            ::||:  |: : | : :   :: : : | |   : :||:          : | |
Db      771 MHAAVTTTHTDWCYKTWMNTWYWDMMTTMBTTTTRNMTTSTNMTNNNNNNMWACTNNNNN 712

Qy      257 TTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGACTTTGTTGCCTATAGA 316
              : |:: :    :   :   : ::|   | :::::    :|::: ||: : :  ||
Db      711 NMWKAYYAHATNNWGCWWNNTDARRTNNTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNN 652

Qy      317 GGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTGTGCTCTACACAGGATG 376
                    :  |      :::|:  :|:   ::::::: |  :::: |   :::
Db      651 NNNNNNNNNNNNNSCCTCTRMMTMRWTMKGDGMTVRKKVKWRDTTCTYVDVWADSWVWWY 592

Qy      377 GATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAGCACGCCTGCCACTGGA 436
              ::: : :      |      :::::    :: :: :|: :::: :  |:  :  :|
Db      591 ANWMRCRDVTYTRNNTYCKSYAHSYWYWSNNAMWYRRYSARNWSSMARWTTRNNWWMSGB 532

Qy      437 GGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAACAGTGCTTTCTCTTAA 496
            ::::   |  :::     : :  :  ::: :     ::   |  ::   :  :: ::
Db      531 VRMRWAGTMWWRHWNNNNTDTRYYWWWKRWARBTTTVYDSMCNAKSMWRGNNWRAMKMWW 472

Qy      497 AAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTGGACATGTGCTTCAGAG 556
            ||  :| |::::|  :   |::   |:::: :   :  : |      | ::  ::::::
```

```
Db        471 AANNDAGAMDHWTYWMGNNTMWMRRAWKMMNMAWCRRAYCCNNNNNRACVWHKHKMWRWT 412

Qy        557 ACCTT---CCTATTTAACCAGGATACAAGTTACATTGTTATGCAATTCCTCTGCTGAGGC 613
              : : :   |    ::|::::: : :: ::||  ::  :| : ::::: :::| :  :
Db        411 WKYMWKAACNNNNBKAMYMRVAWMMYSRDTTNTDWMMWTSDWBWHWYTVDYTMMRAWNNN 352

Qy        614 CCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACAGATTTCGCCCATT 673
                ::         : :::  :|      :|:| :: |  :  | |: : |:  | : |
Db        351 NNNNWRBCKTTSWMWWMDHMNTHCTYGNNTWGSAYBMAAMSMWAAGASNBVTYNWCWRMT 292

Qy        674 ATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTC 733
                    |  ::   : :   :  :| :: : :  :| |:: :: :::|::
Db        291 YMGKTMTNNNNNNKAWYYRTKTVAWCNNRYYYDTAVWTBKRNYKYCYAYBWYYBMYMGKH 232

Qy        734 TTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACA 793
              ::::  : :::::  ::: |:   :  ::::|:::::  : :      :: : : ::
Db        231 HWBWWRRABHRSWNMWWVKCRNKYMVSWHYHAMRYBKWABAVGCNNNWKDRMAHHHWCAT 172

Qy        794 ATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACA 853
                :::::: :::::::: :||:   :   :::: |:::  :: ::::: |: :: |
Db        171 NNNMMWWWYAYMHHMHKKGKAAWTNNKTABRDDHBAHVKTYWYWRYDYWCAMCWMNAKAK 112

Qy        854 AAT 856
              ::|
Db        111 VRT 109


RESULT 2
US-09-248-796A-742/c
; Sequence 742, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 742
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-09-248-796A-742

  Query Match             4.8%;  Score 41.6;  DB 4;  Length 477;
  Best Local Similarity   52.3%;  Pred. No. 0.063;
  Matches   92;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        627 TGAGTTAGGACAGATGCAGCAAGATATTAGGACAGATTTCGCCCATTATTCAGGCTGCAC 686
              |||  |    ||  ||  |   ||||| | || | || |    || | ||   |     |
Db        233 TGATCTTCATCATTTGGTGATGGATATAATGAAATATATTTTTCAGTTTTAGGAAATAAA 174

Qy        687 AACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACT 746
              |  ||| |||  || |  |||||   | | |  |||| ||  |   |||| ||    ||
Db        173 ATAACATAAGCTAAATCTATTTCTCCATGTCTTAATTGTTTTCTAGCTTTTTTAATATCT 114

Qy        747 TTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTT 802
              |||     |   |||   || ||  |||||||| |  |   |||   |  |||||
Db        113 TTTCTAATTCCAGTTTGAGAAATTCTTTCAAATTCTTTACGTAAATTTTTCAATTT 58


RESULT 3
US-09-949-016-15953
; Sequence 15953, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15953
;   LENGTH: 105050
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15953

  Query Match             4.6%;  Score 39.4;  DB 4;  Length 105050;
  Best Local Similarity   48.8%;  Pred. No. 3.1;
  Matches  106;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        639 GATGCAGCAAGATATTAGGACAGATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGA 698
              || || |||  |   || |||          |     | ||| |||  ||||| || | ||
Db      62127 GAGGCTGCAGTAAGCTATGACCACGCCACACTGCATTCCAGCCTGGGCAACAGAGCAAGA 62186

Qy        699 AACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAAC 758
                ||   |   || || |   ||      ||      ||| || |   | |||   |||||
Db      62187 CTCTGACTCAAAAAAAAAAAAAAAGAAAGAAAGAATATTTTTAACATGTCTTATAGATAAC 62246

Qy        759 TTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACT 818
              ||||  ||| |         ||| |      ||    | | | ||||    |  | |
Db      62247 TTTTGAAAAATATCAAATAATATTTTCTTGATCTGACATATTTTTCTCTATACAAATGAA 62306

Qy        819 GGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAA 855
              ||| | |||| ||  ||| || ||||  ||| | |
Db      62307 TTCTAATGAAAAATATGTGGAATATAATGCAAAAACA 62343


RESULT 4
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

  Query Match             4.5%;  Score 38.4;  DB 4;  Length 832;
  Best Local Similarity   10.8%;  Pred. No. 0.67;
  Matches   27;  Conservative  121;  Mismatches  102;  Indels    0;  Gaps    0;

Qy        542 ACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATTC 601
              :: :::|:|:       | |:|:: ::::  ::::  ::  :   :  :  |:  :: : :
Db        253 RMYYKKGSTYWTMKCTCATWCYWYWKYWKRMWSKTCWSGSRGGYMTSYTSTRSYSMYWAS 194

Qy        602 CTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACAG 661
                :| |: :|:  :| : :|:|:  ::  ||:::|:: ::  : ::: ::: :  :
Db        193 WMYTMCWWWGRWWSTYWYMAWGKKWWRYATTWRRAMMWWWAAWTMMWYMWWAWCMSSRGA 134
```

```
Qy        662 ATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAA 721
              |  :   :::: :::: ::::::     | |:|   ::     ::::::  : :: : :
Db        133 AMYRRTMMMWGYRYWWRKKSYRRTRCAWAYAWKTKRSYYWCWRWKWKRCMMMMMMAMAYG 74

Qy        722 TTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTG 781
              :|:  :||::| :: :  : : ::::| :::: : :    :  ::::: :  ::::|:
Db         73 KTMMMRACWKTRYWRWWAWAMWRMWWTMMMMYYWYWRAMKRRWMWRKWRSWSWMWMAWGM 14

Qy        782 TGCAGAGAAA 791
              |: | : :::
Db         13 TRWAARMWWR 4


RESULT 5
US-09-816-095-3
; Sequence 3, Application US/09816095
; Patent No. 6664084
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu
;  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001147
;  CURRENT APPLICATION NUMBER: US/09/816,095
;  CURRENT FILING DATE:  2001-03-26
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 99916
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(99916)
;   OTHER INFORMATION: n = A,T,C or G
US-09-816-095-3

  Query Match             4.4%;  Score 38.2;  DB 4;  Length 99916;
  Best Local Similarity   58.3%;  Pred. No. 6.7;
  Matches   67;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        746 TTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTA 805
              || ||  || ||  ||| | |||||   |  | |||| ||    || |||||  ||  ||
Db      35960 TTATATTCAGAATATTTTCTAATTTATATTTATTTGTTCATTTTAAAAATAATTTTACTA 36019

Qy        806 AAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTTG 860
              || |  ||||| ||   | ||  || ||  | | |||| ||| |      | |||
Db      36020 AATAATATAAAATGCTAATTTTGAAATACTTAGGATTTTCTTACATTTGGTTTTG 36074


RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             4.3%;  Score 37.4;  DB 1;  Length 7218;
  Best Local Similarity   3.4%;  Pred. No. 3.5;
  Matches    5;  Conservative   98;  Mismatches   44;  Indels    0;  Gaps    0;

Qy         15 GGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAAGCTGTTTTCAGTTG 74
              :::::::::::::::::::  ::: ::: : : ::      :::::  :     ::  :
Db       1207 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1148

Qy         75 AGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAATAGCAGGAGCTGGG 134
              ::::   :: ::: ::  :   ::::  ::::::::    : ::: :: :::::  :::
Db       1147 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1088

Qy        135 AAGAAAGGAAGAGGAGGAAGGATCCCA 161
              :::::::::::::::::::::|||  ||
Db       1087 RRRRRRRRRRRRRRRRRRRRRATCGCA 1061


RESULT 7
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             4.3%;  Score 37.2;  DB 4;  Length 640681;
  Best Local Similarity   57.9%;  Pred. No. 30;
  Matches   66;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        746 TTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTA 805
              |||||    ||  ||||| | ||||| ||||||  ||  ||| |||        |||||
```

```
Db     380085 TTTTAAAGGTATATTTTCTATATTTGCTTCAAACAGTTTAGATAAAATCACTCTTTTTTT 380144

Qy        806 AAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTT 859

Db     380145 TTGATTATTATTTTTTTTGTTTAAAGAACTATATTTTTTTTTAAAAAAAATATT 380198


RESULT 8
US-09-949-016-13846/c
; Sequence 13846, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13846
;   LENGTH: 37715
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13846

  Query Match             4.3%;  Score 37;  DB 4;  Length 37715;
  Best Local Similarity   57.3%;  Pred. No. 9.5;
  Matches   67;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        692 AGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTAC 751

Db      35838 AGTAGCAATCTGCATTTCAATAAGACCTAAATAACCCCAAACTCTCCTATGGCATCTTTT 35779

Qy        752 CCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAA 808

Db      35778 ACTAATATGAAGCATAATTTATTTAACTTGAAAAGAATAGGAAAAAAATGTTGACAA 35722


RESULT 9
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             4.3%;  Score 36.8;  DB 4;  Length 1141;
  Best Local Similarity   9.8%;  Pred. No. 2.2;
  Matches   43;  Conservative  183;  Mismatches  212;  Indels    0;  Gaps    0;

Qy        218 TCTGCCTTATAAAAACTTGCTGAATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCC 277

Db         87 WYBCANNTSBRYHARRWKDMKTAYBMTMTNKWGKTGWRHRYWRWRAMBDTVDHHYVTAMN 146

Qy        278 ATAATAGCTTCAAATTATGGTGACTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTT 337

Db        147 NAWTTMCMMDKDDKRTRWWWKKNNNATGWDDDTKYHMWNNNGCBTVTWMVRYKTDRDWSB 206

Qy        338 TGAGATCCTAGAATCTGAGTGTGTGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAG 397

Db        207 KRMNYGMBWWKNWSYDVTYYWWVWDDMCKRKVRRWVRTRGRMRNYMVAWBTAHRRRYNNG 266

Qy        398 TTCCATTTATCCAGAGGCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAA 457

Db        267 WTBAMAYRRWTMNNNNNNAKAMCKRAKYWGWNRABVNSTCTTWKSKTTKVRTSCWANNCR 326

Qy        458 AGAACCTATGTTGCCTAAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTC 517

Db        327 AGDANKDHKWWKWSAAMGVYWNNNNNNNWTYKKARHBARWDWVWHSAWKKWHANAAHYSR 386

Qy        518 TGCCATGAAGCGGAATGTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGAT 577

Db        387 KKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYNNNNNGGRTYYGWTKNKKMWT 446

Qy        578 ACAAGTTACATTGTTATGCAATTCCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGAC 637

Db        447 YYKWKANNCKWRAWDHKTCTHNNTTWWKMKTYWNNCYWKSMTNGKSHRBAAAVYTWYMWW 506

Qy        638 AGATGCAGCAAGATATTA 655

Db        507 WRRYAHANNNNWDYWWKA 524


RESULT 10
US-09-949-016-16179
; Sequence 16179, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16179
;   LENGTH: 157644
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(157644)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16179

  Query Match             4.2%;  Score 36.4;  DB 4;  Length 157644;
  Best Local Similarity   45.7%;  Pred. No. 27;
  Matches  127;  Conservative    0;  Mismatches  151;  Indels    0;  Gaps    0;

Qy        560 TTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATTCCTCTGCTGAGGCCCTGTA 619

Db      39631 TTTTTAATTGTCATTAATTCTGATTACATCAGCAAAAGGAACCCGTTCTGTAGATGGCAA 39690

Qy        620 AAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACAGATTTCGCCCATTATTCAG 679

Db      39691 CTATGTAGATGACATCTCAGAGTCACCAGTTCATTTGTACATTGTTTTAAAATTATTAAG 39750
```

```
Qy        680 GCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCT 739
              | || |   || | |  || ||||   |  ||   | |||| |  |
Db      39751 AGATCTCAGCTGTGATGTTATTTTTATTTGCTAAGTGCTAAACATTAAACATTACTAAAA 39810

Qy        740 ACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAA 799
              |  |   |  || | || |  ||||    || | |  |    ||  | ||  || |
Db      39811 TTTCCATCGGAGAAATTATTTGTTTAAATAATTTTAACAAAGCCTTGATTAGCATAAATA 39870

Qy        800 TTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATG 837
              ||| | ||  || |    || | |||| | |   ||||
Db      39871 TTTCTTAATGGGGTTGCTTGCCAAGTTGAGATACAATG 39908
```

RESULT 11
US-09-949-016-16180
; Sequence 16180, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16180
; LENGTH: 157644
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(157644)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16180

Query Match 4.2%; Score 36.4; DB 4; Length 157644;
Best Local Similarity 45.7%; Pred. No. 27;
Matches 127; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

```
Qy        560 TTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATTCCTCTGCTGAGGCCCTGTA 619
              ||  ||  ||  |    ||  |  ||||||     |        ||  | |||  |     |
Db      39631 TTTTTAATTGTCATTAATTCTGATTACATCAGCAAAAGGAACCCGTTCTGTAGATGGCAA 39690

Qy        620 AAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACAGATTTCGCCCATTATTCAG 679
              | |      |  |   ||||  || ||    ||| | |||   ||      ||||||  ||
Db      39691 CTATGTAGATGACATCTCAGAGTCACCAGTTCATTTGTACATTGTTTTAAAATTATTAAG 39750

Qy        680 GCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCT 739
              | || |   || | |  || ||||   |  ||   | |||| |  |
Db      39751 AGATCTCAGCTGTGATGTTATTTTTATTTGCTAAGTGCTAAACATTAAACATTACTAAAA 39810

Qy        740 ACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAA 799
              |  |   |  || | || |  ||||    || | |  |    ||  | ||  || |
Db      39811 TTTCCATCGGAGAAATTATTTGTTTAAATAATTTTAACAAAGCCTTGATTAGCATAAATA 39870

Qy        800 TTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATG 837
              ||| | ||  || |    || | |||| | |   ||||
Db      39871 TTTCTTAATGGGGTTGCTTGCCAAGTTGAGATACAATG 39908
```

RESULT 12
US-09-949-001-379
; Sequence 379, Application US/09949001
; Patent No. 6825336
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH OSTEOPOROSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL000789
; CURRENT APPLICATION NUMBER: US/09/949,001
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/231,323
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 848
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 379
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-001-379

Query Match 4.2%; Score 36.2; DB 4; Length 601;
Best Local Similarity 55.0%; Pred. No. 2.5;
Matches 71; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy        732 TCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAA 791
              ||||||    ||  || ||||  | |||  ||| ||| || |  || || || | | |||
Db        339 TCTTTTACTTCCTTTTCTACCAGTGACTAGTCCTAATGTGTTCTAACTTTTGAAAATAAA 398

Qy        792 CAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAA 851
              | ||| |   |  |    |  |   ||  ||| | | | ||||     |     ||||||
Db        399 GTACAAATTAGATGGATTTAAACATACGGGAAGTAAGACGGAAATTCATTACTTTTAAAA 458

Qy        852 CAAATCTTG 860
              | |  |||
Db        459 AACAATTTG 467
```

RESULT 13
US-09-949-001-31
; Sequence 31, Application US/09949001
; Patent No. 6825336
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH OSTEOPOROSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL000789
; CURRENT APPLICATION NUMBER: US/09/949,001
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/231,323
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 848
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 312957
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(312957)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-001-31

Query Match 4.2%; Score 36.2; DB 4; Length 312957;
Best Local Similarity 55.0%; Pred. No. 42;
Matches 71; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy        732 TCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAA 791
              ||||||    ||  || ||||  | |||  ||| ||| || |  || || || | | |||
Db      95248 TCTTTTACTTCCTTTTCTACCAGTGACTAGTCCTAATGTGTTCTAACTTTTGAAAATAAA 95307

Qy        792 CAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAA 851
              | ||| |   |  |    |  |   ||  ||| | | | ||||     |     ||||||
Db      95308 GTACAAATTAGATGGATTTAAACATACGGGAAGTAAGACGGAAATTCATTACTTTTAAAA 95367

Qy        852 CAAATCTTG 860
              | |  |||
Db      95368 AACAATTTG 95376
```

```
RESULT 14
US-09-949-001-34
; Sequence 34, Application US/09949001
; Patent No. 6825336
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH OSTEOPOROSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL000789
;  CURRENT APPLICATION NUMBER: US/09/949,001
;  CURRENT FILING DATE:  2003-01-15
;  PRIOR APPLICATION NUMBER: 60/231,323
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34
;   LENGTH: 312972
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(312972)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-001-34

  Query Match             4.2%;  Score 36.2;  DB 4;  Length 312972;
  Best Local Similarity   55.0%;  Pred. No. 42;
  Matches   71;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        732 TCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAA 791
              |||||||     ||  || ||||  | |||  ||| ||| ||  |  || || ||  | | |||
Db      95262 TCTTTTACTTCCTTTTCTACCAGTGACTAGTCCTAATGTGTTCTAACTTTTGAAAATAAA 95321

Qy        792 CAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAA 851
              | ||| |   |  |     |  |   ||  ||| | | | ||||     |    ||||||
Db      95322 GTACAAATTAGATGGATTTAAACATACGGGAAGTAAGACGGAAATTCATTACTTTTAAAA 95381

Qy        852 CAAATCTTG 860
              | |  |||
Db      95382 AACAATTTG 95390


RESULT 15
US-09-949-016-14172/c
; Sequence 14172, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14172
;   LENGTH: 194933
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(194933)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14172

  Query Match             4.2%;  Score 36;  DB 4;  Length 194933;
  Best Local Similarity   51.2%;  Pred. No. 39;
  Matches   84;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        680 GCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCT 739
              || |  |  ||    |||| | || |  |||       ||   |||||  |  ||||
Db     168329 GCAGAGCCTCATTTCAGGACATTTTAAGTCATCGTAGCCTTTTTCTTCCCTTTCTTAAAA 168270

Qy        740 ACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAA 799
              | | |  |||   |||   || || || ||  |||   | ||   | |||   |  | ||
Db     168269 ATCTAACTTTTAACATCTTTTCTCTAAGTTCTGTTTCTAGTGCCTAAAGATCTACCACAA 168210

Qy        800 TTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTT 843
              ||||| ||  | | | |   ||||    | | ||  |  ||||
Db     168209 ATTTTAGAACTGCTCAGCAACCTAGGAGGAGGAAAGAGCAATTT 168166


Search completed: September  1, 2005, 02:32:00
Job time : 207 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2005, 00:19:08 ; Search time 722 Seconds
(without alignments)
7811.833 Million cell updates/sec

Title: US-10-679-865A-1
Perfect score: 862
Sequence: 1 atgatgttgtgtttggagaa.....taattaaaacaaatcttgca 862

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7331713 seqs, 3271544945 residues

Total number of hits satisfying chosen parameters: 14663426

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	862	100.0	862	19	US-10-679-865A-1	Sequence 1, Appli
	2	43.2	5.0	374849	13	US-10-087-192-1627	Sequence 1627, Ap
c	3	42.8	5.0	6103	15	US-10-311-455-1663	Sequence 1663, Ap
c	4	42.8	5.0	8873	17	US-10-221-613-81	Sequence 81, Appl
c	5	42.4	4.9	168749	17	US-10-085-117-250	Sequence 250, App
	6	41.8	4.8	308	18	US-10-424-599-87084	Sequence 87084, A
	7	41.8	4.8	1748	17	US-10-369-493-29582	Sequence 29582, A
c	8	41.6	4.8	674	13	US-10-027-632-203697	Sequence 203697,
c	9	41.6	4.8	674	17	US-10-027-632-203697	Sequence 203697,
c	10	41.4	4.8	428	9	US-09-969-373-1172	Sequence 1172, Ap
c	11	40.4	4.7	6051	21	US-10-741-849-6229	Sequence 6229, Ap
	12	39.4	4.6	96596	11	US-09-997-722-100	Sequence 100, App
	13	39.2	4.5	302	20	US-10-425-115-107288	Sequence 107288,
	14	39.2	4.5	464	19	US-10-433-485A-7	Sequence 7, Appli
c	15	38.8	4.5	93329	19	US-10-322-281-817	Sequence 817, App
c	16	38.8	4.5	3673778	16	US-10-312-841-1	Sequence 1, Appli
	17	38.4	4.5	471	14	US-10-198-846-610	Sequence 610, App
	18	38.4	4.5	579	19	US-10-021-323-13326	Sequence 13326, A
	19	38.2	4.4	99916	9	US-09-816-095-3	Sequence 3, Appli
	20	38.2	4.4	99916	18	US-10-634-905-3	Sequence 3, Appli
c	21	38	4.4	7049	15	US-10-311-455-129	Sequence 129, App
c	22	38	4.4	7049	16	US-10-240-452-5	Sequence 5, Appli
	23	37.8	4.4	416	10	US-09-918-995-36533	Sequence 36533, A
c	24	37.8	4.4	557	10	US-09-814-353-18892	Sequence 18892, A
c	25	37.8	4.4	99116	17	US-10-298-122-1	Sequence 1, Appli
c	26	37.6	4.4	475	18	US-10-424-599-33484	Sequence 33484, A
c	27	37.6	4.4	611	13	US-10-027-632-186042	Sequence 186042,
c	28	37.6	4.4	611	13	US-10-027-632-186043	Sequence 186043,
c	29	37.6	4.4	611	17	US-10-027-632-186042	Sequence 186042,
c	30	37.6	4.4	611	17	US-10-027-632-186043	Sequence 186043,
c	31	37.6	4.4	12968	14	US-10-239-676-202	Sequence 202, App
c	32	37.6	4.4	12968	15	US-10-311-455-2058	Sequence 2058, Ap
c	33	37.6	4.4	12968	15	US-10-240-453-298	Sequence 298, App
c	34	37.6	4.4	12968	18	US-10-221-714A-506	Sequence 506, App
c	35	37.6	4.4	37973	15	US-10-311-455-2170	Sequence 2170, Ap
c	36	37.6	4.4	117329	22	US-10-987-384-1	Sequence 1, Appli
c	37	37.2	4.3	1483	22	US-10-764-212-79	Sequence 79, Appl
c	38	37.2	4.3	5313	15	US-10-311-455-736	Sequence 736, App
c	39	37.2	4.3	5397	15	US-10-311-455-1655	Sequence 1655, Ap
	40	37.2	4.3	640681	9	US-09-790-988-1	Sequence 1, Appli
c	41	37	4.3	546	16	US-10-029-386-7211	Sequence 7211, Ap
c	42	37	4.3	9502	18	US-10-221-714A-456	Sequence 456, App
c	43	37	4.3	17234	19	US-10-433-793-48	Sequence 48, Appl
c	44	37	4.3	53714	18	US-10-052-482-142	Sequence 142, App
	45	37	4.3	133632	13	US-10-087-192-1810	Sequence 1810, Ap

ALIGNMENTS

RESULT 1
US-10-679-865A-1
; Sequence 1, Application US/10679865A
; Publication No. US20040115714A1
; GENERAL INFORMATION:
; APPLICANT: NeoPharm, Inc.
; APPLICANT: Kasid, Usha
; APPLICANT: Gokhale, Prafulla
; APPLICANT: Kumar, Deepak
; APPLICANT: Broustas, Constantinos
; APPLICANT: Ahmad, Imran
; APPLICANT: Dritschilo, Anatoly
; APPLICANT: Rahman, Aquilur
; TITLE OF INVENTION: GENE BRCC-2 AND DIAGNOSITIC AND THERAPEUTIC USES THEREOF
; FILE REFERENCE: 224382
; CURRENT APPLICATION NUMBER: US/10/679,865A
; CURRENT FILING DATE: 2003-10-03
; PRIOR APPLICATION NUMBER: PCT/US02/10848
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/281,796
; PRIOR FILING DATE: 2001-04-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 862
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (294)..(620)

US-10-679-865A-1

Query Match 100.0%; Score 862; DB 19; Length 862;
Best Local Similarity 100.0%; Pred. No. 7.5e-228;
Matches 862; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540

Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720

Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||||
Db        841 TTTAATTAAAACAAATCTTGCA 862
```

RESULT 2
US-10-087-192-1627
; Sequence 1627, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1627
; LENGTH: 374849
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(374849)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1627

Query Match 5.0%; Score 43.2; DB 13; Length 374849;
Best Local Similarity 54.4%; Pred. No. 10;
Matches 87; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy        702 TTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTT 761
              ||  |||  | ||||   || | |||  |   ||    |  || ||||||| |||   ||
Db      12160 TTTCTTTTCAAAAGAATGAAGTTTTAGAATGATTAAAGAGTCAGTTTTACTCATGCATTA 12219

Qy        762 TCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGC 821
              |||||||  ||  || |   | |  ||||||  |||  |     ||| ||     ||
Db      12220 CACAAATTTTCTTGCAAGTTGCAAAAAGAACAATGTAATGATATGTGAGGTTACTTATGC 12279

Qy        822 TAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTTGC 861
              || | ||| |    |     | | ||||||||||  || |
Db      12280 TAATAAAATGATCTTTTAGATGAGTTAAAACAAAAGTTCC 12319
```

RESULT 3
US-10-311-455-1663/c
; Sequence 1663, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1663
; LENGTH: 6103
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1663

Query Match 5.0%; Score 42.8; DB 15; Length 6103;
Best Local Similarity 51.0%; Pred. No. 1.5;
Matches 101; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```
Qy        658 ACAGATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGAC 717
              | |  ||   | |  | ||||   |  | ||| || ||  |||| | ||   ||     |
```

```
Db       2062 AAATTTTAAACTCCATTTTCAAAATCTAAAACTCAAAATAAAACATTATAAAAAATCTCC 2003

Qy        718 CCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAA 777

Db       2002 CAAATAATTCTAAACATTTAAAACAACATCTAAAAAATAACTTTCACTTTTTTTCATAAA 1943

Qy        778 ATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATG 837

Db       1942 ATAATCATTATAAAACAAAATCTTACAAAAAACTCTAATATATAAAATAACAAAAAAATA 1883

Qy        838 GCATTTAATTAAAACAAA 855

Db       1882 AAATAAAATAAAATAAAA 1865


RESULT 4
US-10-221-613-81/c
; Sequence 81, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, ALexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
;  FILE REFERENCE: 5013.1004
;  CURRENT APPLICATION NUMBER: US/10/221,613
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: PCT/EP01/02945
;    DE 10013847.00
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-03-15
;    2000-03-15
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 81
;   LENGTH: 8873
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-81

  Query Match             5.0%;  Score 42.8;  DB 17;  Length 8873;
  Best Local Similarity   52.2%;  Pred. No. 1.8;
  Matches   95;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        674 ATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTC 733

Db       2334 ACTCCGTCTCAAAAAAAAAAAAAACAACTCTATAAAACTAATATAAAAACCATATCTATC 2275

Qy        734 TTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACA 793

Db       2274 TTATTCACCACGAAATACCTAACACATAACAAAATTCCTTACACATAAACACACAATAAA 2215

Qy        794 ATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACA 853

Db       2214 AAATAAATTTTAAACACAATATAATAAAAATAAATAATTAAATATAATTACTATAATAAC 2155

Qy        854 AA 855

Db       2154 AA 2153


RESULT 5
US-10-085-117-250/c
; Sequence 250, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: 529452000121
;  CURRENT APPLICATION NUMBER: US/10/085,117
;  CURRENT FILING DATE:  2002-02-27
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 361
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 250
;   LENGTH: 168749
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: variation
;   LOCATION: (1)...(168749)
;   OTHER INFORMATION: n = any nucleotide
US-10-085-117-250

  Query Match             4.9%;  Score 42.4;  DB 17;  Length 168749;
  Best Local Similarity   52.9%;  Pred. No. 11;
  Matches   91;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        684 CACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCC 743

Db      47051 CAAAACGCACCATAAACAATGTTTAAAAACATTGCAGGCTGGAGAAAATATTTTCATATA 46992

Qy        744 ACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTT 803

Db      46991 ACTTACCCCCAAAATTAATATACAAAACATATAAAGGGTTTTTAAAATCAGAAATATTTT 46932

Qy        804 TAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAA 855

Db      46931 TAAAAAGGAAAAAATGGAAAAAGAATATAAATAGGCAATTGACAGGAAAAGA 46880


RESULT 6
US-10-424-599-87084
; Sequence 87084, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 87084
;   LENGTH: 308
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_49647C.1
US-10-424-599-87084

  Query Match             4.8%;  Score 41.8;  DB 18;  Length 308;
  Best Local Similarity   50.8%;  Pred. No. 0.57;
  Matches  100;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy        664 TTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATT 723

Db         21 TCCGCCCACGCGTCCGATTACAAAGGAAAGGTCTCGTTGTGATATGGTTAAGAAAAATAT 80

Qy        724 CTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTG 783

Db         81 ATTTACCATGTGCAGCACTATCTCTGAACATTTGTTATTACAAAATTATTTCATGTGTCA 140
```

```
Qy        784 CAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTT 843
              | | ||| || |||      |||||  | ||||     |   ||||   ||  |||  |
Db        141 AAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAACAAAAGCAAAT 200

Qy        844 AATTAAAACAAATCTTG 860
              ||  |||||| ||  ||
Db        201 AAAAAAAACATATAATG 217


RESULT 7
US-10-369-493-29582
; Sequence 29582, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 29582
;   LENGTH: 1748
;   TYPE: DNA
;   ORGANISM: Caenorhabditis elegans
US-10-369-493-29582

  Query Match             4.8%;  Score 41.8;  DB 17;  Length 1748;
  Best Local Similarity   53.3%;  Pred. No. 1.4;
  Matches   88;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        693 GAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACC 752
              |||  |  |    |||| |  |    || || ||   | |    | |||  | ||||
Db        638 GAAAAATGCCACTTTTCCAATAATTGCATTTTTTCCTATTAAAATTTACGTATTTTTGAT 697

Qy        753 CATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGA 812
              ||   |||   | ||||| ||||||||    | ||||  || | || |||| ||||
Db        698 AATTTTTTTGGGATATTTGTTTCAAATTTAAAAAAGAATAAAAACAAATTTTCAAAAACC 757

Qy        813 TAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATC 857
              | | |  |||       | ||||  | |||   ||||  ||||
Db        758 AACAATTTCTAAACTTCTGAAAATATAAATTACCAAAAAACAATC 802


RESULT 8
US-10-027-632-203697/c
; Sequence 203697, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 203697
;   LENGTH: 674
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(674)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-203697

  Query Match             4.8%;  Score 41.6;  DB 13;  Length 674;
  Best Local Similarity   54.6%;  Pred. No. 0.99;
  Matches   83;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        706 TTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCA 765
              |||   | |   | ||||| | ||| ||||||  ||  | |||||  |   |   |||
Db        189 TTTAGTTTATGGCTAATTCGTTACAATCTTTTAAACATAATTTTAGTCTGCATACTTCAG 130

Qy        766 AATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGT 825
              ||| |  || |||||    |||   | | | |  | |||  |||||  | | |   |  |
Db        129 AATATTATTAAAATTTGCTAGACTTATATTCAYTTATTTCTAAAGGGAACAATATATTAT 70

Qy        826 TAAAAGTAAATGGCATTTAATTAAAACAAATC 857
              ||  |  |||||  |  | ||  ||   | | |
Db         69 TAGGATAAAATGTCCCTAAAGAAAGTAACAAC 38


RESULT 9
US-10-027-632-203697/c
; Sequence 203697, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 203697
;   LENGTH: 674
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(674)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-203697

  Query Match             4.8%;  Score 41.6;  DB 17;  Length 674;
  Best Local Similarity   54.6%;  Pred. No. 0.99;
  Matches   83;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        706 TTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCA 765
```

```
                ||| | | | ||||| | ||| ||||||  ||  | |||||  |   |   |||
Db        189 TTTAGTTTATGGCTAATTCGTTACAATCTTTTAAACATAATTTTAGTCTGCATACTTCAG 130

Qy        766 AATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGT 825
              ||| |  || |||||    |||   | | | |  | |||  |||||  | | |   | |
Db        129 AATATTATTAAAATTTGCTAGACTTATATTCAYTTATTTCTAAAGGGAACAATATATTAT 70

Qy        826 TAAAAGTAAATGGCATTTAATTAAAACAAATC 857
              ||  |  ||||| |  | ||  ||   | | |
Db         69 TAGGATAAAATGTCCCTAAAGAAAGTAACAAC 38


RESULT 10
US-09-969-373-1172/c
; Sequence 1172, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
;  APPLICANT: Effertz, Roger J.
;  APPLICANT:  Hauge, Brian M.
;  TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
;  FILE REFERENCE: 38-10(52679)A
;  CURRENT APPLICATION NUMBER: US/09/969,373
;  CURRENT FILING DATE:  2001-10-02
;  PRIOR APPLICATION NUMBER: US 09/754,853
;  PRIOR FILING DATE: 2001-01-05
;  PRIOR APPLICATION NUMBER: US 09/760,427
;  PRIOR FILING DATE: 2001-01-13
;  PRIOR APPLICATION NUMBER: US 09/855,768
;  PRIOR FILING DATE: 2001-05-15
;  NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 1172
;   LENGTH: 428
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-969-373-1172

  Query Match             4.8%;  Score 41.4;  DB 9;  Length 428;
  Best Local Similarity   55.1%;  Pred. No. 0.88;
  Matches   81;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy        710 AATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATT 769
              || || |   |  | ||||   | |||  ||   |  | ||   |||   |||    |||
Db        165 AAAAAAATTGATATTTTAATTTTTTTTTATTATATATATATATATATATTATTTTATCATT 106

Qy        770 TGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAA 829
              |  || | ||| | || | ||  || || ||  ||||||   | ||     ||| || ||
Db        105 TATTTAATATTTTTCAAAAAATTAAAAATATGATTAAAATAAAAAATTGATCTAATTCAA 46

Qy        830 AGTAAATGGCATTTAATTAAAACAAAT 856
              | ||  |   |||  || ||| || ||
Db         45 ATTATTTTAAATTAGATCAAATCAGAT 19


RESULT 11
US-10-741-849-6229/c
; Sequence 6229, Application US/10741849
; Publication No. US20050019931A1
; GENERAL INFORMATION:
;  APPLICANT: Roemer, Terry
;  APPLICANT:  Jiang, Bo
;  APPLICANT:  Boone, Charles
;  APPLICANT:  Bussey, Howard
;  TITLE OF INVENTION: Nucleic Acids Encoding Anti-fungal Drug Targets and Methods of
;  TITLE OF INVENTION:  Use
;  FILE REFERENCE: 10182-023-999
;  CURRENT APPLICATION NUMBER: US/10/741,849
;  CURRENT FILING DATE:  2003-12-19
;  PRIOR APPLICATION NUMBER: US 60/434,832
;  PRIOR FILING DATE: 2002-12-19
;  NUMBER OF SEQ ID NOS: 8000
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6229
;   LENGTH: 6051
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-10-741-849-6229

  Query Match             4.7%;  Score 40.4;  DB 21;  Length 6051;
  Best Local Similarity   56.9%;  Pred. No. 6.9;
  Matches   74;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        697 GAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATA 756
              ||  |||  || ||||||  || | | |  |||     | | | |||  ||||  |  ||||
Db       5889 GATTCTTCTTTCCAATAATACTCTAGTGATAATCTCTTGTACAACCAACTTATTACCATT 5830

Qy        757 ACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAA 816
              ||| |||||||| | |||   | ||   | |   | ||||| |    | ||      |
Db       5829 ACTGTTCCAAATATTGTTATTACTGAAAAAATGTAAAATAATACCAATGAACCAATGAGT 5770

Qy        817 CTGGCTAGTT 826
               || ||| ||
Db       5769 ATGACTATTT 5760


RESULT 12
US-09-997-722-100
; Sequence 100, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David
;  APPLICANT:  Engelhard, Eric
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: A-71171/RMS/DCF
;  CURRENT APPLICATION NUMBER: US/09/997,722
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 301
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 100
;   LENGTH: 96596
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-997-722-100

  Query Match             4.6%;  Score 39.4;  DB 11;  Length 96596;
  Best Local Similarity   48.8%;  Pred. No. 57;
  Matches  106;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        639 GATGCAGCAAGATATTAGGACAGATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGA 698
              || || |||  |   || |||          |     | ||| |||  ||||| || | ||
Db      70127 GAGGCTGCAGTAAGCTATGACCACGCCACACTGCATTCCAGCCTGGGCAACAGAGCAAGA 70186

Qy        699 AACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAAC 758
              ||    |   || || |   ||      ||       ||| || |   | |||   |||||
Db      70187 CTCTGACTCAAAAAAAAAAAAAAAGAAAGAAAGAATATTTTTAACATGTCTTATAGATAAC 70246

Qy        759 TTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACT 818
              ||||  ||| |       |||  |      ||    | | | ||||    |   | |
Db      70247 TTTTGAAAAATATCAAATAATATTTTCTTGATCTGACATATTTTTCTCTATACAAATGAA 70306

Qy        819 GGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAA 855
              ||| | |||| ||  ||| || ||||  ||| | |
Db      70307 TTCTAATGAAAAATATGTGGAATATAATGCAAAAACA 70343


RESULT 13
US-10-425-115-107288
; Sequence 107288, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
```

```
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 107288
;   LENGTH: 302
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_29347C.1
US-10-425-115-107288

  Query Match             4.5%;  Score 39.2;  DB 20;  Length 302;
  Best Local Similarity   53.2%;  Pred. No. 3;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        703 TGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTT 762
              || ||||| |  ||    | ||  |   ||||  | |||||  |||| ||    |
Db         52 TGGTTTCATTTTGAAAGCACTCGAAGTTGTCTCCTTTACCCTTTTTTTCCTTATATACGA 111

Qy        763 CCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCT 822
              |   | |||||| || ||    | | ||| || ||||     |||||  | |||
Db        112 CGGCAATTGGTTAAAGTTACATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 171

Qy        823 AGTTAAAAGTAAATGGCATTTAATTAAAACAAATCT 858
              |   ||||  |||    |   ||  |||| ||| ||
Db        172 AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAACT 207


RESULT 14
US-10-433-485A-7
; Sequence 7, Application US/10433485A
; Publication No. US20040131617A1
; GENERAL INFORMATION:
;  APPLICANT: WHITE, Jay A.
;  APPLICANT:  PETKOVICH, P. Martin
;  APPLICANT:  JONES, Glenville
;  APPLICANT:  RAMSHAW, Heather
;  TITLE OF INVENTION: P450RAI-2(P450 Cytochrome 26B), Encoding Nucleic Acid
;  TITLE OF INVENTION:  Molecules and Methods and Uses Thereof
;  FILE REFERENCE: 11812-78
;  CURRENT APPLICATION NUMBER: US/10/433,485A
;  CURRENT FILING DATE:  2003-06-13
;  PRIOR APPLICATION NUMBER: PCT/CA01/01805
;  PRIOR FILING DATE: 2001-12-17
;  PRIOR APPLICATION NUMBER: PCT/CA00/01493
;  PRIOR FILING DATE: 2000-12-15
;  PRIOR APPLICATION NUMBER: 60/178,314
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/171,110
;  PRIOR FILING DATE: 1999-12-16
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
;   LENGTH: 464
;   TYPE: DNA
;   ORGANISM: rattus
US-10-433-485A-7

  Query Match             4.5%;  Score 39.2;  DB 19;  Length 464;
  Best Local Similarity   52.4%;  Pred. No. 3.7;
  Matches   86;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        697 GAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATA 756
              ||||| || || ||||||    |||||| ||     ||     ||  |  |||| |   |
Db        239 GAAACATGTTTACAATAAAGAACAATTCATATTTTACTAAATAACAAATATTTAACAGCA 298

Qy        757 ACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAA 816
              |    |  ||| |   |    ||| || |    ||||| | | |  ||| ||  | | |
Db        299 AAAACTTTAAACTAAATATCTATTTTGAATTATAACAAAATAGTACTTATAATAGTTTAT 358

Qy        817 CTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTTG 860
              |   |  | ||    ||  ||||||   |||| || | |||
Db        359 AAAGACGGACACAAAATTATAACATTTATGAAAAAAAAGTTTTG 402


RESULT 15
US-10-322-281-817/c
; Sequence 817, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc S. Malandro
;  TITLE OF INVENTION: Novel Compositions and Methods in Cancer
;  FILE REFERENCE: 529452001000
;  CURRENT APPLICATION NUMBER: US/10/322,281
;  CURRENT FILING DATE:  2002-12-17
;  NUMBER OF SEQ ID NOS: 866
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 817
;   LENGTH: 93329
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(93329)
;   OTHER INFORMATION: n = A,T,C or G
US-10-322-281-817

  Query Match             4.5%;  Score 38.8;  DB 19;  Length 93329;
  Best Local Similarity   54.1%;  Pred. No. 82;
  Matches   79;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        717 CCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCA 776
              | || ||||  ||| | |  |||     ||||||   |    | | |  ||| ||  | |
Db      46232 CTCAGTTCTAGACATTATGATCTTTAGTCTTTTAAGAAACTGTGTACTTAATATGTATAA 46173

Qy        777 AATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAAT 836
              |||  || | | |   |  | || |   ||||  || |||||  | |   ||| || ||
Db      46172 TATTAAGCTGCGTACTTAATATATATAATAAAAAAATTAACTGCATGGAGCAAAATATAT 46113

Qy        837 GGCATTTAATTAAAACAAATCTTGCA 862
              |  | ||  || |||| | | |
Db      46112 CCTAACTGATAGAATCAAATTTAGAA 46087


Search completed: September  1, 2005, 02:44:08
Job time : 727 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2005, 23:02:12 ; Search time 588 Seconds
(without alignments)
8678.260 Million cell updates/sec

Title: US-10-679-865A-1
Perfect score: 862
Sequence: 1 atgatgttgtgtttggagaa.....taattaaaacaaatcttgca 862

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	862	100.0	862	8	ABS56399	Abs56399 Human BRC
	2	43.4	5.0	2000	8	ADA71938	Ada71938 Rice gene
	3	43.2	5.0	110000	11	ACN44932_0	Acn44932 Mouse gen
c	4	42.8	5.0	6103	6	ABL33690	Abl33690 Human imm
c	5	42.8	5.0	8873	6	ABK31210	Abk31210 Signal tr
c	6	42.8	5.0	8873	6	ABL70173	Abl70173 Chemicall
c	7	42.8	5.0	8873	6	AAS61124	Aas61124 Human gen
c	8	42.6	4.9	2000	8	ADA71938	Ada71938 Rice gene
	9	41.8	4.8	1748	13	ADS51152	Ads51152 Bacterial
c	10	41.4	4.8	2334	10	ADE61946	Ade61946 Human gen
c	11	41.4	4.8	2334	10	ADE60837	Ade60837 Human gen
c	12	41.4	4.8	2334	10	ADE61950	Ade61950 Human gen
c	13	41.4	4.8	22401	4	AAK71911	Aak71911 Human imm
c	14	41.4	4.8	22402	4	AAK71912	Aak71912 Human imm
c	15	40.4	4.7	6051	12	ADP98744	Adp98744 C. albica
	16	39.4	4.6	96596	9	ADA02834	Ada02834 Human MEF
	17	39.4	4.6	96596	10	ADB72572	Adb72572 Human MEF
	18	39.4	4.6	96596	10	ADC85313	Adc85313 Mouse Mef
	19	39.4	4.6	96596	12	ADM74429	Adm74429 Human car
	20	39.2	4.5	464	4	AAH22437	Aah22437 P450RAI-2
	21	39.2	4.5	464	6	ABQ74188	Abq74188 Rat P450R
c	22	38.8	4.5	6160	6	ABK31273	Abk31273 Signal tr
c	23	38.8	4.5	6160	6	ABL70234	Abl70234 Chemicall
c	24	38.8	4.5	93329	13	ABD33597	Abd33597 Murine ca
	25	38.4	4.5	456	4	AAL09259	Aal09259 Human bre
	26	38.4	4.5	471	11	ACN79460	Acn79460 Breast ca
	27	38.4	4.5	579	13	ACN58545	Acn58545 Cotton gy
	28	38.4	4.5	7173	4	ABL16520	Abl16520 Drosophil
	29	38.2	4.4	99916	6	ADI03931	Adi03931 Human enz
	30	38.2	4.4	116297	12	ADQ97587	Adq97587 Human can
c	31	38	4.4	7049	6	ABL32156	Abl32156 Human imm
c	32	38	4.4	7049	6	ABL54305	Abl54305 Chemicall
c	33	38	4.4	110000	3	AAF22305_06	Continuation (7 of
	34	37.8	4.4	416	9	ACH49321	Ach49321 Human leu
c	35	37.8	4.4	557	5	ADL45002	Adl45002 Human ova
c	36	37.6	4.4	12968	4	AAS45494	Aas45494 Chemicall
c	37	37.6	4.4	12968	4	AAS46779	Aas46779 Tumour su
c	38	37.6	4.4	12968	6	ABL34085	Abl34085 Human imm
c	39	37.6	4.4	12968	6	ABK28424	Abk28424 DNA trans
c	40	37.6	4.4	12968	6	ABN80285	Abn80285 Human che
c	41	37.6	4.4	37973	6	ABL34197	Abl34197 Human imm
c	42	37.2	4.3	2354	4	ABL22074	Abl22074 Drosophil
c	43	37.2	4.3	5313	6	ABL32763	Abl32763 Human imm
c	44	37.2	4.3	5397	6	ABL33682	Abl33682 Human imm
	45	37.2	4.3	110000	6	ABA92787_3	Continuation (4 of

ALIGNMENTS

RESULT 1
ABS56399
ID ABS56399 standard; cDNA; 862 BP.
XX
AC ABS56399;
XX
DT 24-JAN-2003 (first entry)
XX
DE Human BRCC2 cDNA.
XX
KW Human; gene; ss; BRCC2; apoptosis; cancer; proliferation; apoptosis;
KW metastasis; inhibitor; ribozyme; antisense; gene therapy; radiotherapy;
KW chemotherapy; anticancer agent; hormone; cell cycle dependent kinase;
KW antisense raf; tumour; therapeutic; Raf-1; chromosome 11q; lung cancer;
KW breast cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 294. .620
FT /*tag= a
FT /product= "BRCC2"
XX
PN WO200281639-A2.
XX
PD 17-OCT-2002.
XX
PF 08-APR-2002; 2002WO-US010848.
XX
PR 06-APR-2001; 2001US-0281796P.
XX
PA (GEOU) UNIV GEORGETOWN.
XX
PI Kasid U, Gokhale P, Kumar D, Broustas C, Ahmad I, Dritschilo A;
PI Rahman A;
XX
DR WPI; 2003-058521/05.
DR P-PSDB; ABG72006.
XX
PT Novel isolated BRCC2 polypeptide which modulates apoptosis and nucleic
PT acid encoding the polypeptide, useful for treating conditions
PT characterized by BRCC2 underexpression, and as diagnostic target for
PT detecting cancer.

XX
PS Claim 2; Fig 1; 48pp; English.
XX
CC The invention discloses an isolated polypeptide, and polynucleotide
CC encoding it, of the human BRCC2 (not defined) protein. The BRCC2
CC polypeptide modulates apoptosis. The polynucleotide, polypeptide and
CC antibodies raised against the polypeptide, can be used for treating or
CC inhibiting cancer cell proliferation, apoptosis and/or metastasis in a
CC cancer, of preferably the breast or lung, patient by administering a
CC small molecule inhibitor of BRCC2, an antibody that specifically binds
CC BRCC2, a ribozyme or antisense oligonucleotide (gene therapy) that
CC modulates BRCC2 expression which can be used in combination with
CC radiotherapy, chemotherapy, biological anticancer agent, hormones or
CC inhibitors of cell cycle dependent kinases. They are also useful for
CC detecting, treating and for evaluating the prognosis of conditions,
CC particularly cancer, characterised by BRCC2 over or underexpression, for
CC identifying compounds that inhibit cancer progression by modulating
CC apoptosis of cancer and as markers for normal physiological function
CC related to cell growth or survival. BRCC2s advantage is that decreasing
CC the amount of BRCC2 induces many of the in vivo effects of antisense raf
CC oligonucleotide, such as tumour growth arrest, tumour regression, tumour
CC cell death and/or potentiate radiation/drug-induced cytotoxicity. In
CC addition, being a therapeutic target potentially downstream of Raf-1,
CC greater specificity is obtained by targeting the action of BRCC2 as
CC compared to Raf-1. The sequence presented is the human BRCC2 cDNA and its
CC gene is located on chromosome 11q
XX
SQ Sequence 862 BP; 272 A; 163 C; 185 G; 242 T; 0 U; 0 Other;

Query Match 100.0%; Score 862; DB 8; Length 862;
Best Local Similarity 100.0%; Pred. No. 1.7e-236;
Matches 862; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGATGTTGTGTTTGGAGAAAAGAGAGAAGAGAATTGAATAAGTACATGACCCCTTGGAA 60

Qy         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGTTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAA 120

Qy        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAGCAGGAGCTGGGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATC 180

Qy        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TAACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGA 240

Qy        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATAAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGA 300

Qy        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTG 360

Qy        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGCTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAG 420

Qy        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAA 480

Qy        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTG 540

Qy        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATT 600

Qy        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACA 660

Qy        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCA 720

Qy        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATT 780

Qy        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCA 840

Qy        841 TTTAATTAAAACAAATCTTGCA 862
              |||||||||||||||||||||||
Db        841 TTTAATTAAAACAAATCTTGCA 862
```

RESULT 2
ADA71938
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 5.0%; Score 43.4; DB 8; Length 2000;
Best Local Similarity 8.2%; Pred. No. 0.13;

```
  Matches   37;  Conservative  224;  Mismatches  185;  Indels    7;  Gaps    1;

Qy        399 TCCATTTATCCAGAGGCAAAAGCACGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAA 458
              |: :::  :|:::::  :::  ::::::|::  :::: ::|   ::  :|  || : :  :
Db        382 TSRRKKRKWCMRKRKYKRMRGYSRMRSCKRARWMKRCRSGRAWKMGCRGCMTCRMKSYGM 441

Qy        459 GAACCTATGTTGCCTAAGGAAACAGTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCT 518
               :: ::     :: :: :  ::  :: : ||:: :|:  : ||: :  :::::    :|
Db        442 MRWKSWKRMASKYKWMSRMYRWRKKKCSRTTMWGKTRGGMMGTMGRCRYKKRSGMKRKCR 501

Qy        519 GCCATGAAGCGGAATGTTCCTGGACATGTGCTTCAGAGACCTTCCTATTTAACCAGGATA 578
              :  : :: :::  ::     :  ::  :::: ::| ::| |::  : :::| : ::: ::
Db        502 RRRWGRMYRMRWKRYYMSARYTMRYCARKKYSYSAARKARCWYRGKGYYWAGMWMKRYKR 561

Qy        579 CAAGTTACATTGTTATGCAATTCCTCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACA 638
              : : : : :::::: : ::  :::::        :|::: : |  :: :: : |  : ::
Db        562 MYMYKMMWWYKRKYSKCSWYCKMSYYA-------SCMKSARKAGAKMCKRSKMSAWSKSM 614

Qy        639 GATGCAGCAAGATATTAGGACAGATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGA 698
              :  :  :  |:      : |::|  |  : ::::: :||  : :|::   |  :  :::
Db        615 RSSRKCRKCASKRSSAKRYAMMGGMTSGSRMSRWKSYTCYWRKWGSMKSTCTWMYYMSKY 674

Qy        699 AACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAAC 758
             ::| :: ::  |: ::: ::::: :: :: :: ::|:  : :   ::   ::| ::
Db        675 TYAKYGSYWRYRYRAWCMYMWRWYYYRYRSYMTYMAWYTSSTRMAMTGMKYSGRYWTSWY 734

Qy        759 TTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACT 818
              :: :|::       ::::: : :::::: | :::: : :   ::: :  : | :: :
Db        735 KYCKCSWKYRSMWYYWSWWWAKTWKMWRRYATRMMWMWYRYSMKWYTWCTMWGYWWYWWR 794

Qy        819 GGCTAGTTAAAAGTAAATGGCATTTAATTAAAA 851
              ::::: :    :  :: :  ::|  : |:::
Db        795 TYMKMRYMWYKCTKTYWYWSATYWTGTWAAWWM 827


RESULT 3
ACN44932_0
WP Sequence split into 4 fragments  LOCUS ACN44932 Accession Acn44932
WP     Fragment Name      Begin     End
WP     ACN44932_0             1    110000
WP     ACN44932_1        100001    210000
WP     ACN44932_2        200001    310000
WP     ACN44932_3        300001    374849
ID   ACN44932 standard; DNA; 374849 BP.
XX
AC   ACN44932;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Mouse genomic sequence mCG5728.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
XX
OS   Mus musculus.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 1627; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 374849 BP; 106870A; 74478C; 74612G; 111149T; 0U; 77400ther;

  Query Match             5.0%;  Score 43.2;  DB 11;  Length 110000;
  Best Local Similarity   54.4%;  Pred. No. 0.86;
  Matches   87;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        702 TTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTT 761
              ||  |||  | ||||   || | |||  |   ||    |  || |||||| |||   ||
Db      12160 TTTCTTTTCAAAAGAATGAAGTTTTAGAATGATTAAAGAGTCAGTTTTACTCATGCATTA 12219

Qy        762 TCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGC 821
              |||||||  ||  || |    | |  ||||||  |||  |     ||| ||      ||
Db      12220 CACAAATTTTCTTGCAAGTTGCAAAAAGAACAATGTAATGATATGTGAGGTTACTTATGC 12279

Qy        822 TAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTTGC 861
              || | ||| |    |     | | ||||||||||  || |
Db      12280 TAATAAAATGATCTTTTAGATGAGTTAAAACAAAAGTTCC 12319


RESULT 4
ABL33690/c
ID   ABL33690 standard; DNA; 6103 BP.
XX
AC   ABL33690;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1663.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
```

```
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1663; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6103 BP; 1794 A; 39 C; 1134 G; 3136 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 42.8;  DB 6;  Length 6103;
  Best Local Similarity   51.0%;  Pred. No. 0.32;
  Matches  101;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy        658 ACAGATTTCGCCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGAC 717
              | |  ||   | |  | ||||   |  | ||| || ||  |||| | ||   ||      |
Db       2062 AAATTTTAAACTCCATTTTCAAAATCTAAAACTCAAAATAAAACATTATAAAAAATCTCC 2003

Qy        718 CCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAA 777
              | |||  ||   |   |||   ||    | | |  |||||||| |   |||  | ||
Db       2002 CAAATAATTCTAAACATTTAAAACAACATCTAAAAAATAACTTTCACTTTTTTTCATAAA 1943

Qy        778 ATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATG 837
              ||  |   | |||  | ||  ||   |||||   |||  |    | | | ||  ||||
Db       1942 ATAATCATTATAAAACAAAATCTTACAAAAAACTCTAATATATAAAATAACAAAAAAATA 1883

Qy        838 GCATTTAATTAAAACAAA 855
                ||  ||| |||  |||
Db       1882 AAATAAAATAAAATAAAA 1865


RESULT 5
ABK31210/c
ID   ABK31210 standard; DNA; 8873 BP.
XX
AC   ABK31210;
XX
DT   23-APR-2002  (first entry)
XX
DE   Signal transduction associated gene modified DNA #27.
XX
KW   Human; signal transduction associated gene; cytosine methylation state;
KW   CpG island; signal transduction associated disease; solid tumour; cancer;
KW   antitumour; cytostatic; mutant; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200200926-A2.
XX
PD   03-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007472.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-147896/19.
XX
PT   Oligonucleotide for diagnosis and therapy of diseases associated with
PT   signal transduction e.g. cancer, comprises chemically modified genomic
PT   sequences of genes associated with signal transduction.
XX
PS   Claim 1; SEQ ID NO 53; 24pp; English.
XX
CC   The present invention relates to chemically modified DNA sequences of
CC   signal transduction associated genes. The DNA sequences are chemically
CC   modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC   Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC   the cytosine methylation state (CpG islands) of these genes, and a method
CC   for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC   genes associated with signal transduction. The genomic DNA can be
CC   obtained from cells or cellular components which contain DNA, e.g. cell
CC   lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC   tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
CC   brain, heart, prostate, lung, breast or liver, histologic object slides,
CC   and all their possible combinations. The sequences of the invention are
CC   useful for the diagnosis and therapy of diseases associated with signal
CC   transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC   chemically pretreated genomic DNA sequences of different genes associated
CC   with signal transduction, or their complementary sequences. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office
XX
SQ   Sequence 8873 BP; 2395 A; 221 C; 2060 G; 4197 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 42.8;  DB 6;  Length 8873;
  Best Local Similarity   52.2%;  Pred. No. 0.38;
  Matches   95;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        674 ATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTC 733
              | || | ||  | || | | ||   ||||  ||   | ||| |   ||  | || |  ||
Db       2334 ACTCCGTCTCAAAAAAAAAAAAAAACAACTCTATAAAACTAATATAAAAACCATATCTATC 2275

Qy        734 TTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACA 793
              || |  |||       |||| |  || |  | |||||   | || ||         | | |
Db       2274 TTATTCACCACGAAATACCTAACACATAACAAAATTCCTTACACATAAACACACAATAAA 2215

Qy        794 ATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACA 853
              | | || ||||||| |  ||| | |    |   | || |||||   |||    ||| |
Db       2214 AAATAAATTTTAAACACAATATAATAAAAATAAATAATTAAATATAATTACTATAATAAC 2155

Qy        854 AA 855
              ||
Db       2154 AA 2153


RESULT 6
ABL70173/c
ID   ABL70173 standard; DNA; 8873 BP.
XX
AC   ABL70173;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Chemically treated cell signalling DNA sequence#32.
XX
KW   Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW   tumour; cytostatic; ds.
XX
OS   Unidentified.
XX
PN   WO200202807-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007471.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154758/20.
```

```
XX
PT   Nucleic acid, useful for diagnosis and therapy of diseases associated
PT   with cell signaling e.g. cancer, comprises chemically modified genomic
PT   sequences of genes associated with cell signaling.
XX
PS   Claim 1; SEQ ID NO 63; 24pp + Sequence Listing; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of at least
CC   18 bases of a segment of chemically pretreated DNA of genes associated
CC   with cell signalling. The activity of the modified sequences of the
CC   invention may be described as cytostatic. The object of the invention is
CC   to provide the chemically modified DNA of genes associated with cell
CC   signalling, as well as oligonucleotides and/or PNA-oligomers for
CC   detecting cytosine methylations, as well as a method which is
CC   particularly suitable for the diagnosis and/or therapy of genetic and
CC   epigenetic parameters of genes associated with cell signalling. The
CC   chemically modified DNA provided by the invention is useful for diagnosis
CC   and therapy of diseases such as solid tumours and cancer. The sequences
CC   given in records ABL70111-ABL70626 represent chemically pre-treated
CC   genomic DNA's of genes associated with cell signalling. Note: The
CC   sequence data for this patent is not represented in the printed
CC   specification, but is based on sequence information supplied by the
CC   European Patent Office
XX
SQ   Sequence 8873 BP; 2395 A; 221 C; 2060 G; 4197 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 42.8;  DB 6;  Length 8873;
  Best Local Similarity   52.2%;  Pred. No. 0.38;
  Matches   95;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        674 ATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTC 733
              | || | ||  | || | | ||   |||| ||   | ||| |   ||  | || |  ||
Db       2334 ACTCCGTCTCAAAAAAAAAAAAAAACAACTCTATAAAACTAATATAAAAACCATATCTATC 2275

Qy        734 TTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACA 793
              || |  |||       |||| |  || | | ||||   | || ||         | | |
Db       2274 TTATTCACCACGAAATACCTAACACATAACAAAATTCCTTACACATAAACACACAATAAA 2215

Qy        794 ATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACA 853
              | | || ||||||| |  ||| | |    |   | || |||||   |||    ||| |
Db       2214 AAATAAATTTTAAACACAATATAATAAAAATAAATAATTAAATATAATTACTATAATAAC 2155

Qy        854 AA 855
              ||
Db       2154 AA 2153


RESULT 7
AAS61124/c
ID   AAS61124 standard; DNA; 8873 BP.
XX
AC   AAS61124;
XX
DT   29-JAN-2002  (first entry)
XX
DE   Human gene regulation-associated gene oligonucleotide #79.
XX
KW   Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW   cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW   asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW   renal disease; Preeclampsia; cardiac allograft vascular disease;
KW   colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW   immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW   nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200177375-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-EP003968.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-017470/02.
XX
PT   New nucleic acid sequences from chemically modified genes associated with
PT   gene regulation, useful for analyzing cytosine methylations for diagnosis
PT   and therapy of diseases e.g. severe combined immunodeficiency disease.
XX
PS   Claim 1; SEQ ID NO 81; 26pp; English.
XX
CC   The invention relates to 224 nucleic acid sequences comprising at least
CC   18 bases of a chemically pretreated gene associated with gene regulation
CC   selected from 43 known genes (or complementary sequences). The chemical
CC   pretreatment converts cytosine bases unmethylated at the 5-position to
CC   uracil or another base with hybridisation behaviour dissimilar to
CC   cytosine, to enable analysis of cytosine methylations. The DNA sequences,
CC   oligomers (or sets/arrays) and method are useful in the diagnosis of
CC   diseases (or predisposition to diseases) associated with gene regulation
CC   and in therapy of such diseases, by enabling analysis of the cytosine
CC   methylation patterns of such genes, kits are provided. They are
CC   especially useful in diagnosis and therapy of e.g. severe combined
CC   immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
CC   and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
CC   syndrome, renal disease, preeclampsia, graft versus-host disease. The
CC   present sequence is a sequence included in the sequence data for this
CC   specification and is associated with the human gene regulation-associated
CC   genes. Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 8873 BP; 2395 A; 221 C; 2060 G; 4197 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 42.8;  DB 6;  Length 8873;
  Best Local Similarity   52.2%;  Pred. No. 0.38;
  Matches   95;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        674 ATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTC 733
              | || | ||  | || | | ||   |||| ||   | ||| |   ||  | || |  ||
Db       2334 ACTCCGTCTCAAAAAAAAAAAAAAACAACTCTATAAAACTAATATAAAAACCATATCTATC 2275

Qy        734 TTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACA 793
              || |  |||       |||| |  || | | ||||   | || ||         | | |
Db       2274 TTATTCACCACGAAATACCTAACACATAACAAAATTCCTTACACATAAACACACAATAAA 2215

Qy        794 ATAAAATTTTTAAAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACA 853
              | | || ||||||| |  ||| | |    |   | || |||||   |||    ||| |
Db       2214 AAATAAATTTTAAACACAATATAATAAAAATAAATAATTAAATATAATTACTATAATAAC 2155

Qy        854 AA 855
              ||
Db       2154 AA 2153


RESULT 8
ADA71938/c
ID   ADA71938 standard; DNA; 2000 BP.
XX
AC   ADA71938;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Rice gene, SEQ ID 5263.
XX
KW   Plant; bacterial infection; fungal infection; viral infection; rice;
KW   gene; ds.
XX
```

```
OS   Oryza sativa.
XX
PN   WO2003000898-A1.
XX
PD   03-JAN-2003.
XX
PF   22-JUN-2001; 2001WO-IB001105.
XX
PR   22-JUN-2001; 2001WO-IB001105.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Chang H,  Chen W,  Cooper B,  Glazebrook J,  Goff SA,  Hou Y;
PI   Katagiri F,  Quan S,  Tao Y,  Whitham S,  Xie Z,  Zhu T,  Zou G;
XX
DR   WPI; 2003-175290/17.
XX
PT   Identifying at least one gene involved in plant resistance or response to
PT   pathogenic infection for conferring resistance or tolerance to a plant to
PT   bacterial, fungal or viral infection by determining or detecting plant
PT   gene expression.
XX
PS   Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC   The present invention relates to a method (M1) for identifying genes
CC   involved in plant resistance or response to pathogenic infection. M1
CC   comprises identifying a gene whose expression is significantly altered in
CC   the incompatible interaction of plant gene expression relative to
CC   expression of the gene in an uninfected plant, in a mutant plant that
CC   does not express a gene associated with response to pathogenic infection,
CC   or in a corresponding incompatible or compatible interaction. (M1) is
CC   useful for conferring resistance to resistance or tolerance to a plant to
CC   bacterial, fungal or viral infection. The present sequence was used to
CC   illustrate the invention.
XX
SQ   Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

  Query Match             4.9%;  Score 42.6;  DB 8;  Length 2000;
  Best Local Similarity   9.6%;  Pred. No. 0.22;
  Matches   59;  Conservative  283;  Mismatches  272;  Indels    3;  Gaps    2;

Qy         65 TTTTCAGTTGAGGATTTTGGTGAGTAGCCACTTGAAATCAAGGGAAATTCCATGAATAGC 124
Db        984 YWAYKWRAYAWSRSRKTWWCTGGKRMATYCGTKMWAAGRWRWRMAWCWYCCMWKWWKMTS 925

Qy        125 AGGAGCTG--GGAAGAAAGGAAGAGGAGGAAGGATCCCACCACTTTTTTCTTCCCATCTA 182
Db        924 CMWWKYWRTWSCWYTMWWGAMRYAYYAMRRRRWTYKWSWRRMYWTMTKWAWTWMTCMCMA 865

Qy        183 ACACCGTGCTTATGCAAAGCATAGTCCATATGAGATCTGCCTTATAAAAACTTGCTGAAT 242
Db        864 KWYMATGWATWMWWRYTMYTYCYAMTCAKCKYKMAMTKWWTTWACAWRATSWRWRAMAGM 805

Qy        243 AAAGCGAATGGAAGTTAAATTTTGGACTTTCTTCCATAATAGCTTCAAATTATGGTGACT 302
Db        804 RWKRYKMKRAYWWRWWRCWKAGWARWMKSRYRWKWKKYATRYYWKMWAMTWWWSWRRWKS 745

Qy        303 TTGTTGCCTATAGAGGGCCAGGAAGTACATTTCTTTGAGATCCTAGAATCTGAGTGTGTG 362
Db        744 YRMWSGMGRMRWSAWRYCSRMKCAKTKYASSARWTKRAKRSYRYRRRWYWKRKGWTYRYR 685

Qy        363 CTCTACACAGGATGGATAGAGCGAGCCTCTGGCAGTTCCATTTATCCAGAGGCAAAAGCA 422
Db        684 YWRSCRMTRARMSKRRKW-AGASMKSCWMYWRGARSMWYSKYSCSAKCCKKTRYMTSSYM 626

Qy        423 CGCCTGCCACTGGAGGCGCTCTTGGGTTCCAACAAAGAACCTATGTTGCCTAAGGAAACA 482
Db        625 STGMYGMYSSYKSMSWTSKMSYMGKMTCTMYTSMKGSTRRSKMGRWSGMSRMYMRWWKKM 566

Qy        483 GTGCTTTCTCTTAAAAGGTACAATCTTGGCTCCTCTGCCATGAAGCGGAATGTTCCTGGA 542
Db        565 RKRKYMRYMKWKCTWRRCMCYRWGYTMYTTSRSRMMYTGRYKARYTSKRRYMWYKYRKYC 506

Qy        543 CATGTGCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATTCC 602
Db        505 WYYYYGMYMKCSYMMRYGYCKACKKCCYAMCWKAAYSGMMMYWYRKYSKWMRMSTKYMWS 446

Qy        603 TCTGCTGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACAGA 662
Db        445 MWYKKCRSMKYGAKGCYGCKMWTYCSYGYMKWYTYMGSYKYSRCYKYMRMYMYKGWMYMM 386

Qy        663 TTTCGCCCATTATTCAG 679
Db        385 YYSAYSSMMTWYYYYAK 369


RESULT 9
ADS51152
ID   ADS51152 standard; cDNA; 1748 BP.
XX
AC   ADS51152;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polynucleotide #5895.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polynucleotide; gene; ss.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 29582; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
```

CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1748 BP; 767 A; 260 C; 200 G; 521 T; 0 U; 0 Other;

Query Match 4.8%; Score 41.8; DB 13; Length 1748;
Best Local Similarity 53.3%; Pred. No. 0.36;
Matches 88; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy        693 GAAGGAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACC 752
              |||  |  |    |||| |  |    || || ||   | |    | |||  | ||||
Db        638 GAAAAATGCCACTTTTCCAATAATTGCATTTTTTCCTATTAAAATTTACGTATTTTTGAT 697

Qy        753 CATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGA 812
                ||   |||   | ||||| ||||||||    | ||||  || | || |||| ||||
Db        698 AATTTTTTTGGGATATTTGTTTCAAATTTAAAAAAGAATAAAAACAAATTTTCAAAAACC 757

Qy        813 TAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATC 857
              | | |  |||       | ||||   | |||   ||||  ||||
Db        758 AACAATTTCTAAACTTCTGAAAATATAAATTACCAAAAAACAATC 802
```

RESULT 10
ADE61946/c
ID ADE61946 standard; DNA; 2334 BP.
XX
AC ADE61946;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human gene NM_000414, SEQ ID NO 7875.
XX
KW Human; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; NM_000414.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2334 BP; 676 A; 435 C; 569 G; 654 T; 0 U; 0 Other;

Query Match 4.8%; Score 41.4; DB 10; Length 2334;
Best Local Similarity 49.8%; Pred. No. 0.53;
Matches 131; Conservative 0; Mismatches 131; Indels 1; Gaps 1;

```
Qy        548 GCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATTCCTCTGC 607
              | |||| || ||||  |  ||||      | ||| ||| | || ||| ||| | ||  |
Db       2334 GGTTCAAAGGCCTTATTGATTAATTTTCCTTCAACTTATAATGCTATCCAAATTCTAGGA 2275

Qy        608 TGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACAGATTTCG 667
              | | ||   ||     |   |  | | | | |    |||| |  |||| |   | |
Db       2274 AACAGGCTTGGGCAAATTTTTTATATGAAGATAAGATCTAAGACAGAAGGAAATCATGTG 2215

Qy        668 CCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTA 727
              ||||   | | |   | | |  | || || |||  | | ||| |   || |||
Db       2214 TAGCTTATAATAGTTCCTGTA-AAAATATGATACATGAAATTAGTAAAAATGAAATCTGC 2156

Qy        728 ACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGA 787
              |  |||  | |   | | ||| | ||  |  | |     |  |  | || |||| |
Db       2155 AGTTTCTGATATCGAAAATGTTAGCAATTTCCCTGCATCTAGTTCTAATACTTTTGCAAA 2096

Qy        788 GAAACAATAAAATTTTTAAAAAG 810
                || ||| | | ||  | || ||
Db       2095 TAATCAACATATTTGGTGAAGAG 2073
```

RESULT 11
ADE60837/c
ID ADE60837 standard; DNA; 2334 BP.
XX
AC ADE60837;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human gene NM_000414, SEQ ID NO 6750.
XX
KW Human; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.

```
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; NM_000414.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC   specification) which encodes one of the polypeptides of the invention
CC   which is differentially expressed during pain. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2334 BP; 676 A; 435 C; 569 G; 654 T; 0 U; 0 Other;

  Query Match             4.8%;  Score 41.4;  DB 10;  Length 2334;
  Best Local Similarity   49.8%;  Pred. No. 0.53;
  Matches  131;  Conservative    0;  Mismatches  131;  Indels    1;  Gaps    1;

Qy        548 GCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATTCCTCTGC 607
              | |||| || ||||  |  ||||     | ||| ||| | || ||| ||| | ||  |
Db       2334 GGTTCAAAGGCCTTATTGATTAATTTTCCTTCAACTTATAATGCTATCCAAATTCTAGGA 2275

Qy        608 TGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACAGATTTCG 667
                | | ||   ||    |   |  | | | | |    |||| |  |||| |   |  |
Db       2274 AACAGGCTTGGGCAAATTTTTTATATGAAGATAAGATCTAAGACAGAAGGAAATCATGTG 2215

Qy        668 CCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTA 727
                ||||    | | |   | | |  | || || |||  | | ||| |   || |||
Db       2214 TAGCTTATAATAGTTCCTGTA-AAAATATGATACATGAAATTAGTAAAAATGAAATCTGC 2156

Qy        728 ACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGA 787
              |   |||  | |    | | ||| | ||  |  | |      | |   | || |||| |
Db       2155 AGTTTCTGATATCGAAAATGTTAGCAATTTCCCTGCATCTAGTTCTAATACTTTTGCAAA 2096

Qy        788 GAAACAATAAAATTTTTAAAAAG 810
               || ||| | | ||  | || ||
Db       2095 TAATCAACATATTTGGTGAAGAG 2073


RESULT 12
ADE61950/c
ID   ADE61950 standard; DNA; 2334 BP.
XX
AC   ADE61950;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human gene NM_000414, SEQ ID NO 7879.
XX
KW   Human; ds; gene; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; NM_000414.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC   specification) which encodes one of the polypeptides of the invention
CC   which is differentially expressed during pain. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2334 BP; 676 A; 435 C; 569 G; 654 T; 0 U; 0 Other;

  Query Match             4.8%;  Score 41.4;  DB 10;  Length 2334;
  Best Local Similarity   49.8%;  Pred. No. 0.53;
  Matches  131;  Conservative    0;  Mismatches  131;  Indels    1;  Gaps    1;
```

```
Qy        548 GCTTCAGAGACCTTCCTATTTAACCAGGATACAAGTTACATTGTTATGCAATTCCTCTGC 607
              | |||| || ||||  |  ||||     | ||| ||| | || ||| ||| | ||  |
Db       2334 GGTTCAAAGGCCTTATTGATTAATTTTCCTTCAACTTATAATGCTATCCAAATTCTAGGA 2275

Qy        608 TGAGGCCCTGTAAAAGGAGTGAGTTAGGACAGATGCAGCAAGATATTAGGACAGATTTCG 667
              | | ||   ||    |   | | | | |    |||| |  |||| |   | |
Db       2274 AACAGGCTTGGGCAAATTTTTTATATGAAGATAAGATCTAAGACAGAAGGAAATCATGTG 2215

Qy        668 CCCATTATTCAGGCTGCACAACACAGAAGGAAACTTGATTTCAATAAGACCCAATTCTTA 727
              ||||    | | |  | | |  | || || |||  | | ||| |   || |||
Db       2214 TAGCTTATAATAGTTCCTGTA-AAAATATGATACATGAAATTAGTAAAAATGAAATCTGC 2156

Qy        728 ACAGTCTTTTCTACCCACTTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGA 787
              |  |||  | |    | | ||| | ||  |  | |      |  |   | || |||| |
Db       2155 AGTTTCTGATATCGAAAATGTTAGCAATTTCCCTGCATCTAGTTCTAATACTTTTGCAAA 2096

Qy        788 GAAACAATAAAATTTTTAAAAAG 810
              || ||| | | ||  | || ||
Db       2095 TAATCAACATATTTGGTGAAGAG 2073
```

RESULT 13
AAK71911/c
ID AAK71911 standard; DNA; 22401 BP.
XX
AC AAK71911;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26723.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.

```
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 26723; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 22401 BP; 7030 A; 3838 C; 4282 G; 7251 T; 0 U; 0 Other;

  Query Match             4.8%;  Score 41.4;  DB 4;  Length 22401;
  Best Local Similarity   60.0%;  Pred. No. 1.4;
  Matches   69;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        746 TTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTA 805
              || | |  | |  | || | |  | |  || || | |     ||||||  ||||||||||
Db      11086 TTCTCCTGAAAGATCTTGCCATCTAGAATCCAAATTTTTTTTGAAACAGCAAAATTTTTA 11027

Qy        806 AAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTTG 860
              ||||||||||| |  |  |||  |||| |    | | | ||  ||||| |||
Db      11026 AAAAGGATAAAACGAAAATATAACTTTAAAAGCAGATCATTAAATTCAAATTTTG 10972


RESULT 14
AAK71912/c
ID   AAK71912 standard; DNA; 22402 BP.
XX
AC   AAK71912;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26724.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
```

PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 26724; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 22402 BP; 7031 A; 3840 C; 4283 G; 7248 T; 0 U; 0 Other;

Query Match 4.8%; Score 41.4; DB 4; Length 22402;
Best Local Similarity 60.0%; Pred. No. 1.4;
Matches 69; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy      746 TTTTACCCATAACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTA 805
            || | |  | |  | || | |  | |  || || | |     ||||||  ||||||||||
Db    11087 TTCTCCTGAAAGATCTTGCCATCTAGAATCCAAATTTTTTTTGAAACAGCAAAATTTTTA 11028

Qy      806 AAAAGGATAAACTGGCTAGTTAAAAGTAAATGGCATTTAATTAAAACAAATCTTG 860
            ||||||||||  |   |  |||  |||| |    | | | ||  ||||| |||
Db    11027 AAAAGGATAAAACGAAAATATAACTTTAAAAGCAGATCATTAAATTCAAATTTTG 10973
```

RESULT 15
ADP98744/c
ID ADP98744 standard; DNA; 6051 BP.
XX
AC ADP98744;
XX
DT 23-SEP-2004 (first entry)
XX
DE C. albicans specific gene, orf19.1084, DNA sequence.
XX
KW Diploid fungal cell; allele; gene disruption cassette;
KW promoter replacement fragment; antifungal; fungicide; gene therapy;
KW infection; Candida albicans; gene; ds.

XX
OS Candida albicans.
XX
PN WO2004056965-A2.
XX
PD 08-JUL-2004.
XX
PF 19-DEC-2003; 2003WO-US040618.
XX
PR 19-DEC-2002; 2002US-0434832P.
XX
PA (ELIT-) ELITRA PHARM INC.
PA (ELIT-) ELITRA CANADA LTD.
XX
PI Roemer T, Jiang B, Boone C, Bussey H;
XX
DR WPI; 2004-500296/47.
DR P-PSDB; ADP99054.
XX
PT Constructing a strain of diploid fungal cells in which both alleles of a
PT gene are modified comprises modifying the alleles of a gene in the fungal
PT cells by recombination using a gene disruption cassette and a promoter
PT replacement fragment.
XX
PS Claim 33; SEQ ID NO 6229; 163pp; English.
XX
CC The invention relates to a novel method for constructing a strain of
CC diploid fungal cells in which both alleles of a gene are modified. The
CC method comprises modifying the alleles of a gene in diploid fungal cells
CC by recombination using a gene disruption cassette and a promoter
CC replacement fragment. The invention further comprises: assembling a
CC collection of diploid fungal cells each of which comprises modified
CC alleles of a different gene; a strain of diploid fungal cells comprising
CC modified alleles of a gene, where the first allele of the gene is
CC inactivated by a gene disruption cassette comprising a nucleotide
CC sequence encoding an expressible selectable marker; and the expression of
CC the second allele of the gene is regulated by a heterologous promoter
CC that is operably linked to the coding region of the second allele of the
CC gene, and where the gene encodes the polypeptide mentioned above; a
CC collection of diploid fungal strains comprising the diploid strains cited
CC above, where substantially all the different genes that encode the above
CC amino acid sequences are modified and are present in different diploid
CC strains in the collection; a nucleic acid molecule microarray comprising
CC nucleic acid molecules, where each nucleic acid molecule comprises a
CC nucleotide sequence that is hybridizable to a target nucleotide sequence
CC comprising any of the 310 nucleotide sequences listed in the
CC specification (ADP98516-ADP98825); identifying a gene that is essential
CC to the survival or growth of a fungus, that contributes to the virulence
CC and/or pathogenicity of a fungus, or that contributes to the resistance
CC of a diploid fungus to an antifungal agent; identifying an antifungal
CC agent that inhibits the growth of a diploid fungus, or a therapeutic
CC agent for treatment of a mammalian disease; correlating changes in the
CC levels of proteins or gene transcripts with the inhibition of growth or
CC proliferation of a diploid fungal cell; a purified or isolated nucleic
CC acid molecule comprising a nucleotide sequence encoding a gene product
CC required for proliferation of Candida albicans, where the gene product
CC consists of any of the above-mentioned amino acid sequences; a vector
CC comprising a promoter operably linked to the nucleic acid molecule cited
CC above; a host cell containing the vector; a purified or isolated
CC polypeptide comprising any of the 61 amino acid sequences given in the
CC specification (ADP96718-ADP96778); a fusion protein comprising a fragment
CC of a first polypeptide fused to a second polypeptide, the fragment
CC consisting of at least 6 consecutive residues of any of ADP98826-ADP99135
CC ; producing a polypeptide; identifying a compound which modulates the
CC activity of a gene product encoded by a nucleic acid comprising any of
CC ADP98516-ADP98825; eliciting an immune response in an animal; a strain of
CC Candida albicans, where a first allele of a gene comprising any of
CC ADP98516-ADP98825 is inactive and a second allele of the gene is under
CC the control of a heterologous promoter; identifying a compound or binding
CC partner that binds to the polypeptide comprising any of ADP98826-
CC ADP99135, or its fragment; identifying a compound having the ability to
CC inhibit growth or proliferation of Candida albicans; inhibiting growth or
CC proliferation of Candida albicans cells; manufacturing an antimycotic
CC compound; treating an infection of a subject by Candida albicans;
CC preventing or containing contamination of an object by Candida albicans,
CC or for preventing or inhibiting formation on a surface of a biofilm
CC comprising Candida albicans; a pharmaceutical composition comprising a
CC therapeutic amount of an agent which reduces the activity or level of a
CC gene product encoded by a nucleic acid comprising any of ADP98516-
CC ADP98825 in a pharmaceutical carrier; an antibody preparation which binds
CC the polypeptide; methods for evaluating a compound against a target gene
CC product encoded by any of ADP98516-ADP98825; identifying an antimycotic
CC compound; a computer or a computer readable medium that comprises at
CC least one of the nucleotide sequences mentioned in the specification or
CC at least one amino acid sequence selected from ADP98826-ADP99135; a
CC method assisted by a computer for identifying a putatively essential gene
CC of a fungus; and a protein array comprising proteins, where at least one
CC protein comprises an amino acid sequence or a portion of an amino acid
CC sequence selected from ADP98516-ADP98825. The novel methods and
CC compositions have fungicide activity. The compositions may be used in
CC gene therapy. The composition and methods are useful for drug screening
CC purposes or for diagnosing, preventing or treating infections associated
CC with Candida albicans. These may also be used for constructing strains
CC useful for identification and validation of gene products as effective
CC targets for therapeutic intervention, for identifying and validating gene
CC products as effective targets for therapeutic intervention, and for
CC collecting identified essential genes. This polynucleotide sequence
CC represents a Candida albicans fungal specific gene of the invention.
CC NOTE: This sequence was downloaded from an electronic sequence listing
CC provided on the WIPO website.
XX
SQ Sequence 6051 BP; 2049 A; 1041 C; 1129 G; 1832 T; 0 U; 0 Other;

Query Match 4.7%; Score 40.4; DB 12; Length 6051;
Best Local Similarity 56.9%; Pred. No. 1.6;
Matches 74; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy 697 GAAACTTGATTTCAATAAGACCCAATTCTTAACAGTCTTTTCTACCCACTTTTACCCATA 756
|| ||| || |||||| || | | | ||| | | | ||| |||| | ||||
Db 5889 GATTCTTCTTTCCAATAATACTCTAGTGATAATCTCTTGTACAACCAACTTATTACCATT 5830

Qy 757 ACTTTTCCAAATTTGGTTCAAATTGTGCAGAGAAACAATAAAATTTTTAAAAAGGATAAA 816
||| |||||||| | ||| | || | | | ||||| | | || |
Db 5829 ACTGTTCCAAATATTGTTATTACTGAAAAAATGTAAAATAATACCAATGAACCAATGAGT 5770

Qy 817 CTGGCTAGTT 826
|| ||| ||
Db 5769 ATGACTATTT 5760

Search completed: September 1, 2005, 00:20:09
Job time : 594 secs